137 File Copy Non Elected Sequences

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:38:47 ; Search time 115.07 Seconds
(without alignments)
31.576 Million cell updates/sec

Title: US-09-446-543A-5
Perfect score: 172
Sequence: 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	104	60.5	117	13	Q9W624	Q9w624 carassius a
2	62	36.0	692	2	Q9L8J6	Q9l8j6 rhodospiril
3	61	35.5	428	6	O97859	O97859 bos taurus
4	57	33.1	54	4	Q9UJF9	Q9ujf9 homo sapien
5	57	33.1	465	4	O60687	O60687 homo sapien
6	56	32.6	428	4	Q9UQ49	Q9uq49 homo sapien
7	56	32.6	629	4	Q9NQE1	Q9nqe1 homo sapien
8	55	32.0	1236	2	Q9JPA4	Q9jpa4 rhodocyclus
9	54.5	31.7	503	2	Q9KRY1	Q9kry1 vibrio chol
10	53.5	31.1	176	3	Q08689	Q08689 saccharomyc
11	53	30.8	376	2	Q9RYP2	Q9ryp2 deinococcus
12	52	30.2	1292	2	Q9ZGE5	Q9zge5 heliobacill
13	51	29.7	294	5	Q19530	Q19530 caenorhabdi
14	51	29.7	315	2	O87474	O87474 burkholderi
15	51	29.7	430	1	O27142	O27142 methanobact
16	51	29.7	863	5	Q9VRV3	Q9vrv3 drosophila
17	50.5	29.4	790	10	Q9M371	Q9m371 arabidopsis
18	50	29.1	503	10	Q9LSC6	Q9lsc6 arabidopsis
19	50	29.1	527	5	O76383	O76383 caenorhabdi
20	49	28.5	128	2	P74747	P74747 synechocyst
21	48.5	28.2	127	5	Q9W2F6	Q9w2f6 drosophila
22	48.5	28.2	232	6	Q9N0A7	Q9n0a7 macaca fasc
23	48.5	28.2	548	10	Q9LFA0	Q9lfa0 arabidopsis
24	48.5	28.2	877	5	Q9NE42	Q9ne42 leishmania
25	48	27.9	223	2	O85605	O85605 prevotella
26	48	27.9	314	2	O06348	O06348 mycobacteri
27	48	27.9	343	3	O74569	O74569 coprinus ci
28	48	27.9	348	5	O45431	O45431 caenorhabdi
29	48	27.9	365	3	Q9UT53	Q9ut53 schizosacch
30	48	27.9	772	10	O81349	O81349 chlamydomon
31	48	27.9	1330	2	Q55284	Q55284 synechocyst
32	48	27.9	1331	2	P73020	P73020 synechocyst
33	48	27.9	1379	10	Q07893	Q07893 antirrhinum
34	48	27.9	1380	10	Q40001	Q40001 hordeum vul
35	48	27.9	1381	10	Q39049	Q39049 arabidopsis
36	48	27.9	1382	10	O22435	O22435 nicotiana t
37	48	27.9	1383	10	O65808	O65808 glycine max
38	47.5	27.6	596	2	O86468	O86468 rhodobacter
39	47	27.3	455	2	O33340	O33340 mycobacteri
40	47	27.3	785	1	O26276	O26276 methanobact
41	47	27.3	940	2	Q9KUW5	Q9kuw5 vibrio chol
42	47	27.3	1084	5	Q9TYW4	Q9tyw4 caenorhabdi
43	47	27.3	1300	11	Q9JM92	Q9jm92 mus musculu
44	47	27.3	1308	11	Q9JM52	Q9jm52 mus musculu
45	46.5	27.0	196	10	Q9LLF7	Q911f7 streptocarp

ALIGNMENTS

RESULT 1
Q9W624
ID Q9W624 PRELIMINARY; PRT; 117 AA.
AC Q9W624;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE C-RF AMIDE PRECURSOR.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Satake H., Minakata H., Fujimoto M.;
RT "Carassius RFamide (C-RF amide).";
RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB020024; BAA76662.1; -.
SQ SEQUENCE 117 AA; 12879 MW; D5DC4CB22038C2B0 CRC64;

Query Match 60.5%; Score 104; DB 13; Length 117;
Best Local Similarity 57.7%; Pred. No. 1.3e-07;
Matches 15; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

```
Qy       6 HSMEIRTPDINPAWYAGRGIRPVGRF 31
           |::: |:|:|:| || |||:||:|||
Db      50 HNVDNRSPEIDPFWYVGRGVRPIGRF 75
```

RESULT 2
Q9L8J6
ID Q9L8J6 PRELIMINARY; PRT; 692 AA.
AC Q9L8J6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BCHH (FRAGMENT).
GN BCHH.
OS Rhodospirillum rubrum.

```
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC   Rhodospirillum.
OX   NCBI_TaxID=1085;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R5;
RX   MEDLINE=20138142; PubMed=10671438;
RA   Cheng Y.S., Brantner C.A., Tsapin A., Collins M.L.P.;
RT   "Role of the H protein in assembly of the photochemical reaction
RT   center and intracytoplasmic membrane in Rhodospirillum rubrum.";
RL   J. Bacteriol. 182:1200-1207(2000).
DR   EMBL; AF202319; AAF37352.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   692 AA;  75453 MW;  96430AE93BF35680 CRC64;


  Query Match             36.0%;  Score 62;  DB 2;  Length 692;
  Best Local Similarity   40.6%;  Pred. No. 0.96;
  Matches   13;  Conservative    4;  Mismatches    9;  Indels    6;  Gaps    1;

Qy       3 AHQHSMEIRTPDINPAWYAG------RGIRPV 28
           | | |:| ||  :|| || |       |:| :
Db     568 AEQVSLETRTRALNPKWYEGMLAHGYEGVRQI 599


RESULT   3
O97859
ID   O97859      PRELIMINARY;      PRT;   428 AA.
AC   O97859;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GANGLIOSIDE SIALIDASE.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99143165; PubMed=9988745;
RA   Miyagi T., Wada T., Iwamatsu A., Hata K., Yoshikawa Y., Tokuyama S.,
RA   Sawada M.;
RT   "Molecular cloning and characterization of a plasma membrane-
RT   associated sialidase specific for gangliosides.";
RL   J. Biol. Chem. 274:5004-5011(1999).
DR   EMBL; AB008184; BAA75071.1; -.
DR   INTERPRO; IPR002860; -.
DR   PFAM; PF02012; BNR; 3.
SQ   SEQUENCE   428 AA;  47916 MW;  418B34F3245A8F21 CRC64;


  Query Match             35.5%;  Score 61;  DB 6;  Length 428;
  Best Local Similarity   40.7%;  Pred. No. 0.8;
  Matches   11;  Conservative    6;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       2 RAHQHSMEIRTPDINPAWYAGRGIRPV 28
           ||  ||: | : |:   |: || |:|:
Db     195 RARPHSLMIYSDDLGATWHHGRLIKPM 221


RESULT   4
Q9UJF9
ID   Q9UJF9      PRELIMINARY;      PRT;    54 AA.
AC   Q9UJF9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DJ479J7.3 (SUSHI-REPEAT PROTEIN (SRPUL)) (FRAGMENT).
GN   DJ479J7.3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lawlor S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035608; CAB55682.1; -.
FT   NON_TER      54     54
SQ   SEQUENCE   54 AA;  6110 MW;  E2F3C39F7B961A9F CRC64;


  Query Match             33.1%;  Score 57;  DB 4;  Length 54;
  Best Local Similarity   56.2%;  Pred. No. 0.34;
  Matches    9;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      12 TPDINPAWYAGRGIRP 27
           || : | |||| |  |
Db      18 TPAVTPTWYAGSGYYP 33


RESULT   5
O60687
ID   O60687      PRELIMINARY;      PRT;   465 AA.
AC   O60687;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   SUSHI-REPEAT PROTEIN.
GN   SRPUL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kurosawa H., Inukai T., Inaba T., Goi K., Chang K.-S., Sinjyo T.,
RA   Rakestraw K.M., Naeve C.W., Look T.A.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF060567; AAC15765.1; -.
DR   INTERPRO; IPR000436; -.
DR   INTERPRO; IPR001128; -.
DR   PFAM; PF00084; sushi; 3.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
SQ   SEQUENCE   465 AA;  52971 MW;  4D752B187FF3EFB8 CRC64;


  Query Match             33.1%;  Score 57;  DB 4;  Length 465;
  Best Local Similarity   56.2%;  Pred. No. 3.3;
  Matches    9;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      12 TPDINPAWYAGRGIRP 27
           || : | |||| |  |
Db      18 TPAVTPTWYAGSGYYP 33


RESULT   6
Q9UQ49
ID   Q9UQ49      PRELIMINARY;      PRT;   428 AA.
AC   Q9UQ49;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   GANGLIOSIDE SIALIDASE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Wada T., Yoshikawa Y., Tokuyama S., Kuwabara M., Akita H., Miyagi T.;
RT   "Cloning, expression, and chromosomal mapping of a human ganglioside
RT   sialidase.";
```

```
RL   Biochem. Biophys. Res. Commun. 261:21-27(1999).
DR   EMBL; AB008185; BAA82611.1; -.
DR   INTERPRO; IPR002860; -.
DR   PFAM; PF02012; BNR; 3.
SQ   SEQUENCE   428 AA;  48252 MW;  35D1DD9359A78C98 CRC64;
```

```
  Query Match             32.6%;  Score 56;  DB 4;  Length 428;
  Best Local Similarity   37.0%;  Pred. No. 4.2;
  Matches   10;  Conservative    6;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 RAHQHSMEIRTPDINPAWYAGRGIRPV 28
           :   ||: | : |:   |: || |||:
Db     195 KTRPHSLMIYSDDLGVTWHHGRLIRPM 221
```

```
RESULT   7
Q9NQE1
ID   Q9NQE1      PRELIMINARY;      PRT;   629 AA.
AC   Q9NQE1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   NURAMINIDASE (EC 3.2.1.18).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Monti E., Bassi M.T., Papini N., Riboni M., Manzoni M., Venerando B.,
RA   Croci G., Preti A., Ballabio A., Tettamanti G., Borsani G.;
RT   "Identification and expression of NEU3, a novel human sialidase
RT   associated to the plasma membrane.";
RL   Biochem. J. 349:343-351(2000).
DR   EMBL; Y18563; CAB96131.1; -.
KW   Hydrolase; Glycosidase.
SQ   SEQUENCE   629 AA;  69702 MW;  97C5464B70E69B4B CRC64;
```

```
  Query Match             32.6%;  Score 56;  DB 4;  Length 629;
  Best Local Similarity   37.0%;  Pred. No. 6.4;
  Matches   10;  Conservative    6;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 RAHQHSMEIRTPDINPAWYAGRGIRPV 28
           :   ||: | : |:   |: || |||:
Db     396 KTRPHSLMIYSDDLGVTWHHGRLIRPM 422
```

```
RESULT   8
Q9JPA4
ID   Q9JPA4      PRELIMINARY;      PRT;  1236 AA.
AC   Q9JPA4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   MG PROTOPORPHYRIN METHYL TRANSFERASE.
GN   BCHH.
OS   Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).
OC   Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC   Rubrivivax.
OX   NCBI_TaxID=28068;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL144;
RA   Nagashima K.V., Shimada K., Matsuura K.;
RT   "Phylogenetic analysis of photosynthetic genes of Rhodocyclus
RT   gelatinosus: Possibility of horizontal gene transfer in purple
RT   bacteria.";
RL   Photosyn. Res. 36:185-191(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL144;
RX   MEDLINE=94132007; PubMed=8300574;
RA   Nagashima K.V., Matsuura K., Ohyama S., Shimada K.;
RT   "Primary structure and transcription of genes encoding B870 and
RT   photosynthetic reaction center apoproteins from Rubrivivax
RT   gelatinosus.";
RL   J. Biol. Chem. 269:2477-2484(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL144;
RA   Igarashi N., Shimada K., Matsuura K., Nagashima K.V.;
RT   "PHOTOSYNTHETIC GENE CLUSTER IN PURPLE BACTERIUM, RUBRIVIVAX
RT   GELATINOSUS.";
RL   (in) Garab G. (eds.);
RL   Photosynthesis. Mechanisms and Effects IV:2889-2892;
RL   Kluwer Academic Publishing (1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL144;
RA   Menin L., Yoshida M., Jaquinod M., Nagashima K.V., Matsuura K.,
RA   Parot P., Vermeglio A.;
RT   "Dark aerobic growth conditions induce the synthesis of a high
RT   midpoint potential cytochrome c8 in the photosynthetic bacterium
RT   Rubrivivax gelatinosus.";
RL   Biochemistry 0:0-0(1999).
DR   EMBL; AB034704; BAA94057.1; -.
KW   Transferase.
SQ   SEQUENCE   1236 AA;  134729 MW;  84051C045638520C CRC64;
```

```
  Query Match             32.0%;  Score 55;  DB 2;  Length 1236;
  Best Local Similarity   34.4%;  Pred. No. 18;
  Matches   11;  Conservative    6;  Mismatches    9;  Indels    6;  Gaps    1;

Qy       3 AHQHSMEIRTPDINPAWYAG------RGIRPV 28
           : | ::| ||  :|| || |       |:| :
Db    1112 SEQVALETRTRMLNPKWYEGMLEHGYEGVRQI 1143
```

```
RESULT   9
Q9KRY1
ID   Q9KRY1      PRELIMINARY;      PRT;   503 AA.
AC   Q9KRY1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   SUN/NUCLEOLAR PROTEIN FAMILY PROTEIN.
GN   VC1502.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004228; AAF94657.1; -.
DR   TIGR; VC1502; -.
SQ   SEQUENCE   503 AA;  55997 MW;  2ABD94A2356C9E48 CRC64;
```

```
  Query Match             31.7%;  Score 54.5;  DB 2;  Length 503;
  Best Local Similarity   36.7%;  Pred. No. 8.2;
  Matches   11;  Conservative    6;  Mismatches   12;  Indels    1;  Gaps    1;
```

```
Qy       1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGR 30
           |  :  ||:|: |      |: || :|| |:
Db     435 SSSASHSVELDTTQAR-EWFMGRDVRPEGQ 463
```

```
RESULT  10
Q08689
ID   Q08689      PRELIMINARY;      PRT;   176 AA.
AC   Q08689;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CHROMOSOME XV READING FRAME ORF YOR253W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97298311; PubMed=9153759;
RA   Jauniaux J.C., Poirey R.;
RT   "Sequencing analysis of a 36.8 kb fragment of yeast chromosome XV
RT   reveals 26 open reading frames including SEC63, CDC31, SUG2, GCD1,
RT   RBL2, PNT1, PAC1 and VPH1.";
RL   Yeast 13:483-487(1997).
DR   EMBL; Z75161; CAA99475.1; -.
DR   INTERPRO; IPR000182; -.
DR   INTERPRO; IPR000345; -.
DR   PFAM; PF00583; Acetyltransf; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
SQ   SEQUENCE   176 AA;  19727 MW;  4F09DC597A690BA0 CRC64;
```

```
  Query Match             31.1%;  Score 53.5;  DB 3;  Length 176;
  Best Local Similarity   31.2%;  Pred. No. 3.8;
  Matches   10;  Conservative    5;  Mismatches   14;  Indels    3;  Gaps    1;

Qy       1 SRAHQHSMEIRTP---DINPAWYAGRGIRPVG 29
           |  |||:: :  |   |:   |:   |   ||
Db     120 SECHQHNVFVYLPAVDDLTKQWFIAHGFEQVG 151
```

```
RESULT  11
Q9RYP2
ID   Q9RYP2      PRELIMINARY;      PRT;   376 AA.
AC   Q9RYP2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ADENINE DEAMINASE-RELATED PROTEIN.
GN   DRA0268.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001863; AAF12376.1; -.
DR   TIGR; DRA0268; -.
SQ   SEQUENCE   376 AA;  39845 MW;  7AB7FF32F8C45651 CRC64;
```

```
  Query Match             30.8%;  Score 53;  DB 2;  Length 376;
  Best Local Similarity   40.0%;  Pred. No. 9.9;
  Matches   12;  Conservative    6;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGR 30
           :||  |: ::  || : |  ||:  || ||
Db     177 ARALAHAPDVSRPDRHRACGAGQRRRPAGR 206
```

```
RESULT  12
Q9ZGE5
ID   Q9ZGE5      PRELIMINARY;      PRT;  1292 AA.
AC   Q9ZGE5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MG CHELATASE SUBUNIT H BCHH.
GN   BCHH.
OS   Heliobacillus mobilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Heliobacterium group; Heliobacillus.
OX   NCBI_TaxID=28064;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99061957; PubMed=9843979;
RA   Xiong J., Inoue K., Bauer C.E.;
RT   "Tracking molecular evolution of photosynthesis by characterization of
RT   a major photosynthesis gene cluster from Heliobacillus mobilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
DR   EMBL; AF080002; AAC84033.1; -.
SQ   SEQUENCE   1292 AA;  144853 MW;  323AA0517B07448D CRC64;
```

```
  Query Match             30.2%;  Score 52;  DB 2;  Length 1292;
  Best Local Similarity   38.5%;  Pred. No. 51;
  Matches   10;  Conservative    3;  Mismatches    7;  Indels    6;  Gaps    1;

Qy       9 EIRTPDINPAWYAG------RGIRPV 28
           | ||  :|| || |       |:| :
Db    1176 ETRTKTLNPKWYEGMLKHGYEGVREI 1201
```

```
RESULT  13
Q19530
ID   Q19530      PRELIMINARY;      PRT;   294 AA.
AC   Q19530; Q19606;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   F19H6.1 PROTEIN.
GN   F19H6.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mcmurray A.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z50873; CAA90762.1; -.
DR   EMBL; Z68115; CAA90762.1; JOINED.
DR   EMBL; Z68115; CAA92169.1; -.
DR   EMBL; Z50873; CAA92169.1; JOINED.
DR   HSSP; P24941; 1CKP.
DR   INTERPRO; IPR000719; -.
DR   INTERPRO; IPR001245; -.
DR   INTERPRO; IPR002290; -.
DR   PFAM; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
```

```
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
SQ   SEQUENCE   294 AA;  34006 MW;  071D46644D049F43 CRC64;
```

Query Match 29.7%; Score 51; DB 5; Length 294;
Best Local Similarity 44.8%; Pred. No. 15;
Matches 13; Conservative 2; Mismatches 12; Indels 2; Gaps 1;

```
Qy       3 AHQHSMEIRTPDINPA--WYAGRGIRPVG 29
           || ||  |   || ||  :  | ||  :|
Db     139 AHMHSKRIMHRDIKPANVFITGNGIVKLG 167
```

```
RESULT  14
O87474
ID   O87474      PRELIMINARY;      PRT;   315 AA.
AC   O87474;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   TRIHYDROXYTOLUENE OXYGENASE.
GN   DNTD.
OS   Burkholderia cepacia (Pseudomonas cepacia).
OC   Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC   Burkholderia.
OX   NCBI_TaxID=292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DNT;
RX   MEDLINE=93194809; PubMed=8449889;
RA   Suen W.C., Spain J.C.;
RT   "Cloning and characterization of Pseudomonas sp. strain DNT genes for
RT   2,4-dinitrotoluene degradation.";
RL   J. Bacteriol. 175:1831-1837(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DNT;
RA   Haigler B.E., Johnson G.R., Suen W.C., Spain J.C.;
RT   "2,4,5-Trihydroxytoluene oxygenase catalyzes meta-ring cleavage of
RT   2,4,5-trihydroxytoluene: Biochemical and genetic evidence.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF076848; AAD12738.1; -.
DR   INTERPRO; IPR000486; -.
DR   PFAM; PF01013; Extradiol_dioxy; 1.
DR   PRODOM; PD000977; -; 1.
SQ   SEQUENCE   315 AA;  34745 MW;  E99261179022961E CRC64;
```

Query Match 29.7%; Score 51; DB 2; Length 315;
Best Local Similarity 41.7%; Pred. No. 16;
Matches 15; Conservative 3; Mismatches 12; Indels 6; Gaps 2;

```
Qy       1 SRAHQHSMEIRTPD---INPAWYAGRGIRP---VGR 30
           || |  | ||:  |   :   | | || :|   |||
Db     216 SRMHHCSFEIQDLDAQFLGNKWLASRGWKPGWGVGR 251
```

```
RESULT  15
O27142
ID   O27142      PRELIMINARY;      PRT;   430 AA.
AC   O27142;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CONSERVED PROTEIN.
GN   MTH1070.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanobacterium.
OX   NCBI_TaxID=2166;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000877; AAB85559.1; -.
DR   INTERPRO; IPR002510; -.
DR   PFAM; PF01523; PmbA_TldD; 1.
SQ   SEQUENCE   430 AA;  46062 MW;  C4FAE47C111749E6 CRC64;
```

Query Match 29.7%; Score 51; DB 1; Length 430;
Best Local Similarity 69.2%; Pred. No. 22;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      14 DINPAWYAGRGIR 26
           |||| | |||  |
Db     191 DINPEWVAGRACR 203
```

Search completed: April 17, 2001, 15:48:03
Job time: 556 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:39:52 ; Search time 39.1 Seconds
(without alignments)
27.159 Million cell updates/sec

Title: US-09-446-543A-5
Perfect score: 172
Sequence: 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	172	100.0	98	1	PRRP_BOVIN	P81264 bos taurus
2	163	94.8	83	1	PRRP_RAT	P81278 rattus norv
3	158	91.9	87	1	PRRP_HUMAN	P81277 homo sapien
4	51	29.7	798	1	UNR_RAT	P18395 rattus norv
5	51	29.7	962	1	UVRA_METTH	O26543 methanobact
6	50	29.1	973	1	UVRA_RHIME	P56899 rhizobium m
7	48.5	28.2	239	1	6PGL_SYNY3	P74618 synechocyst
8	48.5	28.2	1882	1	Y468_MYCPN	P75109 mycoplasma
9	48	27.9	424	1	SAMB_SALTY	P23832 salmonella
10	48	27.9	960	1	UVRA_TREPA	O83527 treponema p
11	47.5	27.6	1083	1	T2D3_HUMAN	O00268 homo sapien
12	47	27.3	406	1	DHI2_RABIT	P51976 oryctolagus
13	47	27.3	972	1	UVRA_MYCTU	P94972 mycobacteri
14	46.5	27.0	770	1	AVP3_ARATH	P31414 arabidopsis
15	46	26.7	342	1	Y762_METJA	Q58172 methanococc
16	46	26.7	347	1	Y576_METJA	Q57996 methanococc
17	46	26.7	510	1	YCGB_ECOLI	P29013 escherichia
18	46	26.7	798	1	UNR_HUMAN	O75534 homo sapien
19	46	26.7	836	1	TGLK_RABIT	P22758 oryctolagus
20	46	26.7	943	1	UVRA_HAEIN	P44410 haemophilus
21	45.5	26.5	272	1	TRA2_DROVI	O02008 drosophila
22	45.5	26.5	374	1	YHHJ_ECOLI	P31993 escherichia
23	45.5	26.5	411	1	PUR7_ARATH	P38025 arabidopsis
24	45.5	26.5	704	1	LIPA_NEIMB	Q05013 neisseria m
25	45	26.2	482	1	R167_YEAST	P39743 saccharomyc
26	45	26.2	676	1	EXL1_HUMAN	Q92935 homo sapien
27	45	26.2	817	1	TGLK_HUMAN	P22735 homo sapien
28	45	26.2	824	1	TGLK_RAT	P23606 rattus norv
29	45	26.2	940	1	UVRA_ECOLI	P07671 escherichia
30	45	26.2	941	1	UVRA_SALTY	P37434 salmonella
31	45	26.2	953	1	UVRA_RICPR	Q9zcc3 rickettsia
32	45	26.2	1033	1	ACAC_ARATH	Q91y77 arabidopsis
33	44.5	25.9	152	1	GST3_HUMAN	O14880 homo sapien
34	44.5	25.9	241	1	YG3D_YEAST	P53281 saccharomyc
35	44.5	25.9	407	1	Y1I6_MYCTU	P72052 mycobacteri
36	44.5	25.9	474	1	ATZA_PSESD	P72156 pseudomonas
37	44.5	25.9	777	1	BISC_ECOLI	P20099 escherichia
38	44.5	25.9	1783	1	Y468_MYCGE	Q49460 mycoplasma
39	44	25.6	179	1	RK6_GUITH	O46908 guillardia
40	44	25.6	264	1	Y355_BUCAI	P57436 buchnera ap
41	44	25.6	358	1	PLCA_ERWCH	Q47499 erwinia chr
42	44	25.6	383	1	CYCR_CHRVI	O82947 chromatium
43	44	25.6	952	1	UVRA_THETH	Q56242 thermus aqu
44	44	25.6	970	1	UVRA_SYNY3	P73412 synechocyst
45	44	25.6	1014	1	UVRA_STRCO	Q9z507 streptomyce

ALIGNMENTS

RESULT 1
PRRP_BOVIN
ID PRRP_BOVIN STANDARD; PRT; 98 AA.
AC P81264;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE RELEASING PEPTIDE PRRP20].
GN PRH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.
RC TISSUE=Brain;
RX MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RL Nature 393:272-276(1998).
CC -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; AB015417; BAA29025.1; -.
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 22
FT PEPTIDE 23 53 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT PEPTIDE 33 53 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 53 53 AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ SEQUENCE 98 AA; 10544 MW; 08AC35A13B0FA908 CRC64;

Query Match 100.0%; Score 172; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 2.5e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
|||||||||||||||||||||||||||||||
Db 23 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 53

```
RESULT   2
PRRP_RAT
ID   PRRP_RAT       STANDARD;      PRT;    83 AA.
AC   P81278;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015418; BAA29026.1; -.
KW   Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     21       BY SIMILARITY.
FT   PEPTIDE      22     52       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      33     52       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      52     52       AMIDATION (G-53 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   83 AA;  9215 MW;  D0C75A264EEE4F29 CRC64;


  Query Match             94.8%;  Score 163;  DB 1;  Length 83;
  Best Local Similarity   93.5%;  Pred. No. 4.3e-17;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          ||||||||| |||||||||| ||||||||||
Db     22 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 52


RESULT   3
PRRP_HUMAN
ID   PRRP_HUMAN     STANDARD;      PRT;    87 AA.
AC   P81277;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015419; BAA29027.1; -.
DR   MIM; 602663; -.
KW   Hormone; Amidation; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   PEPTIDE      23     53       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      34     53       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      53     53       AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   87 AA;  9639 MW;  229A2F3F50CF981B CRC64;


  Query Match             91.9%;  Score 158;  DB 1;  Length 87;
  Best Local Similarity   90.3%;  Pred. No. 2.4e-16;
  Matches   28;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          || |:||||||||||||||| |||||||||
Db     23 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 53


RESULT   4
UNR_RAT
ID   UNR_RAT        STANDARD;      PRT;   798 AA.
AC   P18395;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UNR PROTEIN.
GN   UNR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=90370473; PubMed=2204029;
RA   Jeffers M., Paciucci R., Pellicer A.;
RT   "Characterization of unr; a gene closely linked to N-ras.";
RL   Nucleic Acids Res. 18:4891-4899(1990).
CC   -!- FUNCTION: RNA-BINDING PROTEIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52311; CAA36549.1; -.
DR   PIR; S11210; S11210.
DR   HSSP; P15277; 1MJC.
```

```
DR   InterPro; IPR002059; -.
DR   Pfam; PF00313; CSD; 8.
DR   PROSITE; PS00352; COLD_SHOCK; 4.
KW   RNA-binding; Repeat.
FT   DOMAIN       26     87       CSD 1.
FT   DOMAIN      136    179       CSD 2 (INCOMPLETE).
FT   DOMAIN      186    245       CSD 3.
FT   DOMAIN      297    337       CSD 4 (INCOMPLETE).
FT   DOMAIN      349    410       CSD 5.
FT   DOMAIN      447    507       CSD 6.
FT   DOMAIN      519    579       CSD 7.
FT   DOMAIN      610    670       CSD 8.
FT   DOMAIN      674    735       CSD 9.
SQ   SEQUENCE   798 AA;  88894 MW;  F484B3FA8B0995A4 CRC64;


  Query Match             29.7%;  Score 51;  DB 1;  Length 798;
  Best Local Similarity   43.5%;  Pred. No. 9.2;
  Matches   10;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       6 HSMEIRTPDINPAWYAGRGIRPV 28
           ||:   | : ||  |:|: |||:
Db     583 HSVNGITEEANPTIYSGKVIRPL 605


RESULT   5
UVRA_METTH
ID   UVRA_METTH     STANDARD;      PRT;   962 AA.
AC   O26543;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR MTH443.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000828; AAB84949.1; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      38     45       ATP (POTENTIAL).
FT   NP_BIND     649    656       ATP (POTENTIAL).
FT   ZN_FING     748    774       C4-TYPE.
SQ   SEQUENCE   962 AA;  108395 MW;  2C0EF7FC41CCD060 CRC64;


  Query Match             29.7%;  Score 51;  DB 1;  Length 962;
  Best Local Similarity   39.5%;  Pred. No. 11;
  Matches   15;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYAG----------------RGIRPVGRF 31
             |||  ||| | |                || || |||
Db     703 RTPRSNPATYTGVFTHIRELFAQTPEARKRGYRP-GRF 739


RESULT   6
UVRA_RHIME
ID   UVRA_RHIME     STANDARD;      PRT;   973 AA.
AC   P56899;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RA   Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
RA   Boistard P., Gouzy J., Kahn D., Thebault P., Goffeau A.,
RA   Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
RA   Vandenbol M., Puehler A., Becker A., Weidner S.;
RL   Submitted (MAY-2000) to the SWISS-PROT data bank.
RN   [2]
RP   SEQUENCE OF 1-140 FROM N.A.
RC   STRAIN=2021;
RX   MEDLINE=99430868; PubMed=10503543;
RA   Tapias A., Barbe J.;
RT   "Regulation of divergent transcription from the uvrA-ssb promoters in
RT   Sinorhizobium meliloti.";
RL   Mol. Gen. Genet. 262:121-130(1999).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF125162; AAF03210.1; -.
DR   InterPro; IPR001617; -.
DR   PROSITE; PS00211; ABC_TRANSPORTER; PARTIAL.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      34     41       ATP (POTENTIAL).
FT   NP_BIND     662    669       ATP (POTENTIAL).
FT   ZN_FING     761    787       C4-TYPE.
```

```
FT   CONFLICT     19      19       G -> A (IN REF. 2).
FT   CONFLICT     67      67       F -> S (IN REF. 2).
SQ   SEQUENCE   973 AA;  107191 MW;  3E1A8B14527A47FE CRC64;


  Query Match             29.1%;  Score 50;  DB 1;  Length 973;
  Best Local Similarity   39.5%;  Pred. No. 16;
  Matches   15;  Conservative    2;  Mismatches    3;  Indels   18;  Gaps    3;

Qy      11 RTPDINPA----------WYAG------RGIRPVGRF 31
           |||  |||          |:||      || :| |||
Db     716 RTPRSNPATYTGAFTPIRDWFAGLPEAKARGYQP-GRF 752


RESULT   7
6PGL_SYNY3
ID   6PGL_SYNY3     STANDARD;      PRT;   239 AA.
AC   P74618;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   6-PHOSPHOGLUCONOLACTONASE (EC 3.1.1.31) (6PGL).
GN   PGL OR DEVB OR SLL1479.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA   Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA   Yamada M., Yasuda M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
RN   [2]
RP   SEQUENCE OF 1-19.
RX   MEDLINE=97443974; PubMed=9298645;
RA   Sazuka T., Ohara O.;
RT   "Towards a proteome project of cyanobacterium Synechocystis sp.
RT   strain PCC6803: linking 130 protein spots with their respective
RT   genes.";
RL   Electrophoresis 18:1252-1258(1997).
CC   -!- FUNCTION: HYDROLYSIS OF 6-PHOSPHOGLUCONOLACTONE TO 6-
CC       PHOSPHOGLUCONATE.
CC   -!- CATALYTIC ACTIVITY: 6-PHOSPHO-D-GLUCONO-1,5-LACTONE + H(2)O = 6-
CC       PHOSPHO-D-GLUCONATE.
CC   -!- PATHWAY: SECOND STEP IN PENTOSE PHOSPHATE PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE GLUCOSAMINE/GALACTOSAMINE-6-PHOSPHATE
CC       ISOMERASE FAMILY. 6-PHOSPHOGLUCONOLACTONASE SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90916; BAA18726.1; ALT_INIT.
DR   InterPro; IPR000457; -.
DR   Pfam; PF01182; Glucosamine_iso; 1.
KW   Hydrolase.
FT   INIT_MET      0       0
SQ   SEQUENCE   239 AA;  26351 MW;  9C64A0A342325917 CRC64;


  Query Match             28.2%;  Score 48.5;  DB 1;  Length 239;
  Best Local Similarity   42.3%;  Pred. No. 6.1;
  Matches   11;  Conservative    5;  Mismatches    9;  Indels    1;  Gaps    1;

Qy       5 QHSM-EIRTPDINPAWYAGRGIRPVG 29
           ||:: ||  |: :|  |  | |:| |
Db     198 QHALGEIFAPEADPQQYPARFIQPQG 223


RESULT   8
Y468_MYCPN
ID   Y468_MYCPN     STANDARD;      PRT;  1882 AA.
AC   P75109; Q50317;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN MG468 HOMOLOG (K05_ORF1882).
GN   MPN684 OR MP158.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
RN   [2]
RP   SEQUENCE OF 1-1848 FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=96177562; PubMed=8604303;
RA   Hilbert H., Himmelreich R., Plagens H., Herrmann R.;
RT   "Sequence analysis of 56 kb from the genome of the bacterium
RT   Mycoplasma pneumoniae comprising the dnaA region, the atp operon and a
RT   cluster of ribosomal protein genes.";
RL   Nucleic Acids Res. 24:628-639(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: SOME, TO MG064.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000017; AAB95806.1; -.
DR   EMBL; U34816; AAC43650.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     16     36       POTENTIAL.
FT   TRANSMEM    987   1007       POTENTIAL.
FT   TRANSMEM   1037   1057       POTENTIAL.
FT   TRANSMEM   1080   1100       POTENTIAL.
FT   TRANSMEM   1154   1174       POTENTIAL.
FT   TRANSMEM   1759   1779       POTENTIAL.
FT   TRANSMEM   1807   1827       POTENTIAL.
FT   TRANSMEM   1828   1848       POTENTIAL.
FT   TRANSMEM   1851   1871       POTENTIAL.
SQ   SEQUENCE   1882 AA;  209442 MW;  03CFA4D99A7120ED CRC64;


  Query Match             28.2%;  Score 48.5;  DB 1;  Length 1882;
  Best Local Similarity   32.3%;  Pred. No. 52;
  Matches   10;  Conservative    6;  Mismatches   12;  Indels    3;  Gaps    1;

Qy       1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
             :|   : : ::||     |||   |:|  ||
Db    1188 NRNFNYRLNLQTPTEQSGWYA---IQPYSRF 1215


RESULT   9
```

```
SAMB_SALTY
ID   SAMB_SALTY     STANDARD;      PRT;   424 AA.
AC   P23832;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   SAMB PROTEIN.
GN   SAMB.
OS   Salmonella typhimurium.
OG   Plasmid 60-MDa cryptic.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=91123176; PubMed=1991707;
RA   Nohmi T., Hakura A., Nakai Y., Watanabe M., Murayama S.Y.,
RA   Sofuni T.;
RT   "Salmonella typhimurium has two homologous but different umuDC
RT   operons: cloning of a new umuDC-like operon (samAB) present in a
RT   60-megadalton cryptic plasmid of S. typhimurium.";
RL   J. Bacteriol. 173:1051-1063(1991).
CC   -!- FUNCTION: INVOLVED IN UV PROTECTION AND MUTATION.
CC   -!- SIMILARITY: BELONGS TO THE IMPB/MUCB/SAMB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90202; BAA14226.1; -.
DR   PIR; B38176; B38176.
DR   InterPro; IPR001126; -.
DR   Pfam; PF00817; IMS; 1.
KW   Plasmid; SOS mutagenesis; DNA repair.
SQ   SEQUENCE   424 AA;  47727 MW;  FF8C47476CC58A2B CRC64;


  Query Match           27.9%;  Score 48;  DB 1;  Length 424;
  Best Local Similarity   50.0%;  Pred. No. 13;
  Matches    9;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      10 IRTPDINPAWYAGRGIRP 27
           |  |     |:||||| |
Db     381 INHPGKGKIWFAGRGIAP 398


RESULT  10
UVRA_TREPA
ID   UVRA_TREPA     STANDARD;      PRT;   960 AA.
AC   O83527;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR TP0514.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NICHOLS;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001227; AAC65502.1; -.
DR   TIGR; TP0514; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      35     42       ATP (POTENTIAL).
FT   NP_BIND     657    664       ATP (POTENTIAL).
FT   ZN_FING     270    297       C4-TYPE.
FT   ZN_FING     756    782       C4-TYPE.
SQ   SEQUENCE   960 AA;  106010 MW;  32F78624B19F7ABF CRC64;


  Query Match           27.9%;  Score 48;  DB 1;  Length 960;
  Best Local Similarity   36.8%;  Pred. No. 31;
  Matches   14;  Conservative    1;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYAG----------------RGIRPVGRF 31
           |||  ||| | |                || :| |||
Db     711 RTPRSNPATYVGVFTDIRMLFSQVPEAKMRGYKP-GRF 747


RESULT  11
T2D3_HUMAN
ID   T2D3_HUMAN     STANDARD;      PRT;  1083 AA.
AC   O00268; Q99721;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TRANSCRIPTION INITIATION FACTOR TFIID 135 KDA SUBUNIT (TAFII-135)
DE   (TAFII135) (TAFII-130) (TAFII130).
GN   TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97336072; PubMed=9192867;
RA   Mengus G., May M., Carre L., Chambon P., Davidson I.;
RT   "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
RT   of the retinoic acid, vitamin D3, and thyroid hormone receptors in
RT   mammalian cells.";
RL   Genes Dev. 11:1381-1395(1997).
RN   [2]
RP   SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=97098442; PubMed=8942982;
RA   Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT   "Molecular cloning and analysis of two subunits of the human TFIID
RT   complex: hTAFII130 and hTAFII100.";
```

```
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
CC   -!- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
CC       PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
CC       ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
CC       BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
CC   -!- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC       NUMBER OF TBP-ASSOCIATED FACTORS (TAFS).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y11354; CAA72189.1; -.
DR   EMBL; U75308; AAC50901.1; -.
DR   HSSP; P19656; 1AFH.
DR   MIM; 601796; -.
KW   Transcription regulation; Nuclear protein.
FT   DOMAIN       39     42       POLY-HIS.
FT   DOMAIN       52     57       POLY-ALA.
FT   DOMAIN       98    101       POLY-GLY.
FT   DOMAIN      142    148       POLY-ALA.
FT   DOMAIN      268    275       POLY-PRO.
FT   DOMAIN      331    337       POLY-ALA.
FT   DOMAIN      680    683       POLY-PRO.
FT   DOMAIN      808    813       POLY-ALA.
FT   DOMAIN      828    831       POLY-ASP.
FT   CONFLICT    105    117       PGPPSPRRPLVPA -> GRGLLQQRGGRES
FT                                (IN REF. 2).
FT   CONFLICT    233    264       MISSING (IN REF. 2).
FT   CONFLICT    293    293       P -> L (IN REF. 2).
SQ   SEQUENCE   1083 AA;  109943 MW;  A6453827572A0752 CRC64;


  Query Match             27.6%;  Score 47.5;  DB 1;  Length 1083;
  Best Local Similarity   30.8%;  Pred. No. 41;
  Matches   12;  Conservative    4;  Mismatches    8;  Indels   15;  Gaps    1;

Qy       1 SRAHQHSMEIRTPDI---------------NPAWYAGRG 24
           | || | :  |||::               :||  || |
Db      36 SAAHHHHLAPRTPEVRAAAAGALGNHVVSGSPAGAAGAG 74


RESULT  12
DHI2_RABIT
ID   DHI2_RABIT     STANDARD;      PRT;   406 AA.
AC   P51976;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CORTICOSTEROID 11-BETA-DEHYDROGENASE, ISOZYME 2 (EC 1.1.1.146) (11-
DE   DH2) (11-BETA-HYDROXYSTEROID DEHYDROGENASE 2) (11-BETA-HSD2) (NAD-
DE   DEPENDENT 11-BETA-HYDROXYSTEROID DEHYDROGENASE).
GN   HSD11B2.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=Kidney cortex;
RX   MEDLINE=95269702; PubMed=7750480;
RA   Naray-Fejes-Toth A., Fejes-Toth G.;
RT   "Expression cloning of the aldosterone target cell-specific 11 beta-
RT   hydroxysteroid dehydrogenase from rabbit collecting duct cells.";
RL   Endocrinology 136:2579-2586(1995).
CC   -!- FUNCTION: HAS A ROLE IN MODULATING GLUCOCORTICOID ACTIVITY BOTH AT
CC       THE LEVEL OF THE MINERALOCORTICOID RECEPTOR AND THE GLUCOCORTICOID
CC       RECEPTOR. USES NADH WHILE 11-DH1 USES NADPH. CATALYZES NON
CC       REVERSIBLY THE CONVERSION OF CORTISOL TO THE INACTIVE METABOLITE
CC       CORTISONE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: A 11-BETA-HYDROXYSTEROID + NAD(+) = A
CC       11-OXOSTEROID + NADH.
CC   -!- SUBCELLULAR LOCATION: MICROSOMAL.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U23835; AAA86387.1; -.
DR   HSSP; P14061; 1FDV.
DR   InterPro; IPR002198; -.
DR   Pfam; PF00106; adh_short; 1.
DR   PROSITE; PS00061; ADH_SHORT; 1.
KW   Oxidoreductase; NAD; Microsome.
FT   NP_BIND      82    111       NAD (BY SIMILARITY).
FT   ACT_SITE    232    232       BY SIMILARITY.
SQ   SEQUENCE   406 AA;  44181 MW;  F14579663918425D CRC64;


  Query Match             27.3%;  Score 47;  DB 1;  Length 406;
  Best Local Similarity   25.0%;  Pred. No. 18;
  Matches   10;  Conservative    6;  Mismatches    8;  Indels   16;  Gaps    1;

Qy       2 RAHQHSMEIRTPDINPA---------------WYAGRGI 25
           |   ||: :  ||::|                 :| |||:
Db     305 REFLHSLRLALPDLSPVVDAITDALLAARPRPRYYPGRGL 344


RESULT  13
UVRA_MYCTU
ID   UVRA_MYCTU     STANDARD;      PRT;   972 AA.
AC   P94972;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR RV1638 OR MTCY06H11.02.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
```

```
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z85982; CAB06633.1; -.
DR   TubercuList; Rv1638; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      32     39       ATP (POTENTIAL).
FT   NP_BIND     654    661       ATP (POTENTIAL).
FT   ZN_FING     257    285       C4-TYPE (ATYPICAL).
FT   ZN_FING     753    779       C4-TYPE.
SQ   SEQUENCE   972 AA;  106131 MW;  8937A764E592D981 CRC64;


  Query Match             27.3%;  Score 47;  DB 1;  Length 972;
  Best Local Similarity   36.8%;  Pred. No. 43;
  Matches   14;  Conservative    1;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYAG----------------RGIRPVGRF 31
           |||  ||| | |                 || :| |||
Db     708 RTPRSNPATYTGVFDKIRTLFAATTEAKVRGYQP-GRF 744


RESULT  14
AVP3_ARATH
ID   AVP3_ARATH     STANDARD;      PRT;   770 AA.
AC   P31414; Q41919;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PYROPHOSPHATE-ENERGIZED VACUOLAR MEMBRANE PROTON PUMP (EC 3.6.1.1)
DE   (PYROPHOSPHATE-ENERGIZED INORGANIC PYROPHOSPHATASE) (H+-PPASE).
GN   AVP-3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=92179265; PubMed=1311852;
RA   Sarafian V., Kim Y., Poole R.J., Rea P.A.;
RT   "Molecular cloning and sequence of cDNA encoding the pyrophosphate-
RT   energized vacuolar membrane proton pump of Arabidopsis thaliana.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:1775-1779(1992).
RN   [2]
RP   SEQUENCE OF 554-671 FROM N.A.
RC   STRAIN=CV. COLUMBIA; TISSUE=Green siliques;
RA   Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
RL   Submitted (OCT-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CONTRIBUTES TO THE TRANSTONOPLAST (FROM CYTOSOL TO
CC       VACUOLE LUMEN) H+-ELECTROCHEMICAL POTENTIAL DIFFERENCE. THE
CC       H+-PPASE ESTABLISHES A H+ GRADIENT OF SIMILAR AND OFTEN GREATER
CC       MAGNITUDE THAN THE H+-ATPASE ON THE SAME MEMBRANE.
CC   -!- CATALYTIC ACTIVITY: PYROPHOSPHATE + H(2)O = 2 ORTHOPHOSPHATE.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. VACUOLAR
CC       (TONOPLAST).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81892; AAA32754.1; -.
DR   EMBL; Z17694; CAA79038.1; -.
DR   PIR; A38230; A38230.
KW   Hydrogen ion transport; Hydrolase; Transmembrane; Magnesium.
FT   DOMAIN       1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    14     34       POTENTIAL.
FT   DOMAIN      35    100       INTRAVACUOLAR (POTENTIAL).
FT   TRANSMEM   101    121       POTENTIAL.
FT   DOMAIN     122    142       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   143    163       POTENTIAL.
FT   DOMAIN     164    195       INTRAVACUOLAR (POTENTIAL).
FT   TRANSMEM   196    216       POTENTIAL.
FT   DOMAIN     217    327       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   328    348       POTENTIAL.
FT   DOMAIN     349    368       INTRAVACUOLAR (POTENTIAL).
FT   TRANSMEM   369    389       POTENTIAL.
FT   DOMAIN     390    405       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   406    426       POTENTIAL.
FT   DOMAIN     427    484       INTRAVACUOLAR (POTENTIAL).
FT   TRANSMEM   485    505       POTENTIAL.
FT   DOMAIN     506    542       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   543    563       POTENTIAL.
FT   DOMAIN     564    573       INTRAVACUOLAR (POTENTIAL).
FT   TRANSMEM   574    594       POTENTIAL.
FT   DOMAIN     595    645       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   646    666       POTENTIAL.
FT   DOMAIN     667    667       INTRAVACUOLAR (POTENTIAL).
FT   TRANSMEM   668    688       POTENTIAL.
FT   DOMAIN     689    746       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   747    767       POTENTIAL.
FT   DOMAIN     768    770       INTRAVACUOLAR (POTENTIAL).
FT   CONFLICT   576    576       T -> A (IN REF. 2).
FT   CONFLICT   584    584       L -> P (IN REF. 2).
SQ   SEQUENCE   770 AA;  80819 MW;  CE7132B42B299860 CRC64;


  Query Match             27.0%;  Score 46.5;  DB 1;  Length 770;
  Best Local Similarity   26.3%;  Pred. No. 40;
  Matches   10;  Conservative    9;  Mismatches   10;  Indels    9;  Gaps    1;

Qy       1 SRAHQHSMEIRTPDI---------NPAWYAGRGIRPVG 29
           |||  |:::: || :          | |::   :: ||
Db     565 SRAGIHTVDVLTPKVIIGLLVGAMLPYWFSAMTMKSVG 602


RESULT  15
Y762_METJA
ID   Y762_METJA     STANDARD;      PRT;   342 AA.
AC   Q58172;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0762.
GN   MJ0762.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
```

```
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: STRONG, TO M.JANNASCHII MJ0576 AND TO S.POMBE MALATE
CC       PERMEASE (MAE1).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67521; AAB98753.1; -.
DR   TIGR; MJ0762; -.
KW   Hypothetical protein; Transmembrane; Transport.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     39     59       POTENTIAL.
FT   TRANSMEM     79     99       POTENTIAL.
FT   TRANSMEM    108    128       POTENTIAL.
FT   TRANSMEM    142    162       POTENTIAL.
FT   TRANSMEM    175    195       POTENTIAL.
FT   TRANSMEM    207    227       POTENTIAL.
FT   TRANSMEM    242    262       POTENTIAL.
FT   TRANSMEM    276    296       POTENTIAL.
FT   TRANSMEM    304    324       POTENTIAL.
SQ   SEQUENCE   342 AA;  39534 MW;  08EFEC3E2C4955D8 CRC64;


  Query Match             26.7%;  Score 46;  DB 1;  Length 342;
  Best Local Similarity   43.5%;  Pred. No. 21;
  Matches   10;  Conservative    2;  Mismatches    7;  Indels    4;  Gaps    1;

Qy       7 SMEIRTPDINPAWYAGRGIRPVG 29
           |  |:   :|| ||    | |||
Db     131 SESIKLDHVNPGWY----IPPVG 149


Search completed: April 17, 2001, 15:48:45
Job time: 533 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:34:31 ; Search time 61.54 Seconds
(without alignments)
9.677 Million cell updates/sec

Title: US-09-446-543A-5
Perfect score: 172
Sequence: 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	172	100.0	31	3	US-09-105-678A-7	Sequence 7, Appli
2	172	100.0	31	3	US-09-105-678A-31	Sequence 31, Appl
3	172	100.0	32	3	US-09-105-678A-32	Sequence 32, Appl
4	172	100.0	33	3	US-09-105-678A-33	Sequence 33, Appl
5	163	94.8	31	3	US-09-105-678A-8	Sequence 8, Appli
6	163	94.8	31	3	US-09-105-678A-37	Sequence 37, Appl
7	163	94.8	31	4	US-09-172-353-4	Sequence 4, Appli
8	163	94.8	32	3	US-09-105-678A-38	Sequence 38, Appl
9	163	94.8	33	3	US-09-105-678A-39	Sequence 39, Appl
10	161	93.6	29	3	US-09-105-678A-29	Sequence 29, Appl
11	158	91.9	31	3	US-09-105-678A-9	Sequence 9, Appli
12	158	91.9	31	3	US-09-105-678A-43	Sequence 43, Appl
13	158	91.9	32	3	US-09-105-678A-44	Sequence 44, Appl
14	158	91.9	33	3	US-09-105-678A-45	Sequence 45, Appl
15	115	66.9	20	3	US-09-105-678A-34	Sequence 34, Appl
16	115	66.9	21	3	US-09-105-678A-35	Sequence 35, Appl
17	115	66.9	22	3	US-09-105-678A-36	Sequence 36, Appl
18	111	64.5	20	3	US-09-105-678A-40	Sequence 40, Appl
19	111	64.5	21	3	US-09-105-678A-41	Sequence 41, Appl
20	111	64.5	22	3	US-09-105-678A-42	Sequence 42, Appl
21	109	63.4	19	3	US-09-105-678A-30	Sequence 30, Appl
22	109	63.4	20	3	US-09-105-678A-46	Sequence 46, Appl
23	109	63.4	21	3	US-09-105-678A-47	Sequence 47, Appl
24	109	63.4	22	3	US-09-105-678A-48	Sequence 48, Appl
25	104	60.5	21	3	US-09-105-678A-28	Sequence 28, Appl
26	51	29.7	239	2	US-08-712-709-7	Sequence 7, Appli
27	51	29.7	239	2	US-09-049-671-3	Sequence 3, Appli
28	51	29.7	239	3	US-09-295-068-3	Sequence 3, Appli
29	51	29.7	239	3	US-09-111-444-7	Sequence 7, Appli
30	50	29.1	349	1	US-08-118-270-71	Sequence 71, Appl
31	50	29.1	349	5	PCT-US93-08528-71	Sequence 71, Appl
32	46	26.7	555	2	US-08-982-232-7	Sequence 7, Appli
33	46	26.7	555	2	US-08-982-232-14	Sequence 14, Appl
34	46	26.7	774	3	US-08-902-632-2	Sequence 2, Appli
35	46	26.7	774	3	US-09-073-354-1	Sequence 1, Appli
36	46	26.7	774	3	US-08-656-005A-1	Sequence 1, Appli
37	46	26.7	774	4	US-09-073-259-1	Sequence 1, Appli
38	46	26.7	774	4	US-09-363-095-1	Sequence 1, Appli
39	45	26.2	440	3	US-08-985-335-3	Sequence 3, Appli
40	45	26.2	940	4	US-09-078-347A-1	Sequence 1, Appli
41	44.5	25.9	152	2	US-08-959-865-1	Sequence 1, Appli
42	44	25.6	162	4	US-08-992-176-9	Sequence 9, Appli
43	44	25.6	4551	4	US-09-320-878-1	Sequence 1, Appli
44	43	25.0	386	2	US-08-663-310-6	Sequence 6, Appli
45	43	25.0	386	2	US-09-006-491-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT   1
US-09-105-678A-7
; Sequence 7, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-7


  Query Match             100.0%;  Score 172;  DB 3;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31


RESULT   2
US-09-105-678A-31
; Sequence 31, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-31


  Query Match             100.0%;  Score 172;  DB 3;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31


RESULT   3
US-09-105-678A-32
; Sequence 32, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-32


  Query Match             100.0%;  Score 172;  DB 3;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31


RESULT   4
US-09-105-678A-33
; Sequence 33, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
```

```
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-33


  Query Match             100.0%;  Score 172;  DB 3;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31


RESULT   5
US-09-105-678A-8
; Sequence 8, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-8


  Query Match             94.8%;  Score 163;  DB 3;  Length 31;
  Best Local Similarity   93.5%;  Pred. No. 1.6e-18;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          ||||||||| |||||||||| |||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT   6
US-09-105-678A-37
; Sequence 37, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-37


  Query Match             94.8%;  Score 163;  DB 3;  Length 31;
  Best Local Similarity   93.5%;  Pred. No. 1.6e-18;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          ||||||||| |||||||||| |||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT   7
US-09-172-353-4
; Sequence 4, Application US/09172353
; Patent No. 6197530
; GENERAL INFORMATION:
;  APPLICANT: Stricker-Kongra, Alain
;  APPLICANT: Gu, Wei
;  TITLE OF INVENTION: GPR10 AS A TARGET FOR IDENTIFYING WEIGHT MODULATING COMPOUNDS
;  FILE REFERENCE: 07334/102001
;  CURRENT APPLICATION NUMBER: US/09/172,353
;  CURRENT FILING DATE: 1998-10-14
;  NUMBER OF SEQ ID NOS: 7
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 31
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-172-353-4


  Query Match             94.8%;  Score 163;  DB 4;  Length 31;
  Best Local Similarity   93.5%;  Pred. No. 1.6e-18;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          ||||||||| |||||||||| ||||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT   8
US-09-105-678A-38
; Sequence 38, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-38


  Query Match             94.8%;  Score 163;  DB 3;  Length 32;
  Best Local Similarity   93.5%;  Pred. No. 1.6e-18;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          ||||||||| |||||||||| ||||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT   9
US-09-105-678A-39
; Sequence 39, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-39


  Query Match             94.8%;  Score 163;  DB 3;  Length 33;
  Best Local Similarity   93.5%;  Pred. No. 1.7e-18;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          ||||||||| |||||||||| ||||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT  10
US-09-105-678A-29
; Sequence 29, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-29


  Query Match             93.6%;  Score 161;  DB 3;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 3e-18;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVG 29
          |||||||||||||||||||||||||||||
Db      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVG 29


RESULT  11
US-09-105-678A-9
; Sequence 9, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-9


  Query Match             91.9%;  Score 158;  DB 3;  Length 31;
  Best Local Similarity   90.3%;  Pred. No. 9.2e-18;
  Matches   28;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          || |:|||||||||||||||| |||||||||
Db      1 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 31


RESULT  12
US-09-105-678A-43
; Sequence 43, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  43:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-43


  Query Match             91.9%;  Score 158;  DB 3;  Length 31;
  Best Local Similarity   90.3%;  Pred. No. 9.2e-18;
  Matches   28;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          || |:|||||||||||||||| |||||||||
Db      1 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 31
```

```
RESULT  13
US-09-105-678A-44
; Sequence 44, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-44


  Query Match             91.9%;  Score 158;  DB 3;  Length 32;
  Best Local Similarity   90.3%;  Pred. No. 9.6e-18;
  Matches   28;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          || |:|||||||||||||||| |||||||||
Db      1 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 31


RESULT  14
US-09-105-678A-45
; Sequence 45, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-45


  Query Match             91.9%;  Score 158;  DB 3;  Length 33;
  Best Local Similarity   90.3%;  Pred. No. 1e-17;
  Matches   28;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          || |:|||||||||||||||| |||||||||
Db      1 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 31


RESULT  15
US-09-105-678A-34
; Sequence 34, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
```

; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-105-678A-34

Query Match 66.9%; Score 115; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      12 TPDINPAWYAGRGIRPVGRF 31
           ||||||||||||||||||||
Db       1 TPDINPAWYAGRGIRPVGRF 20
```

Search completed: April 17, 2001, 15:39:47
Job time: 316 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:32:53 ; Search time 116.94 Seconds
(without alignments)
15.154 Million cell updates/sec

Title: US-09-446-543A-5
Perfect score: 172
Sequence: 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	172	100.0	31	18	W31371	Bovine G protein-c
2	172	100.0	31	20	W97218	Bovine pituitary-d
3	172	100.0	31	20	W87613	Bovine 19P2 ligand
4	172	100.0	31	20	W95188	Bovine pituitary-d
5	172	100.0	31	21	B10347	Bovine oxytocin se
6	172	100.0	31	21	Y49290	19P2 ligand peptid
7	172	100.0	31	21	Y49298	19P2 ligand peptid
8	172	100.0	32	18	W31372	Bovine G protein-c
9	172	100.0	32	20	W95189	Bovine pituitary-d
10	172	100.0	32	21	B10348	Bovine oxytocin se
11	172	100.0	33	18	W31373	Bovine G protein-c
12	172	100.0	33	20	W95190	Bovine pituitary-d
13	172	100.0	33	21	B10349	Bovine oxytocin se
14	172	100.0	33	21	Y49297	19P2 ligand peptid
15	172	100.0	98	18	W31382	Bovine genome deri
16	172	100.0	98	18	W31368	Bovine G protein-c
17	172	100.0	98	20	W97224	Bovine genome-driv
18	172	100.0	98	20	W97217	Bovine pituitary-d
19	172	100.0	98	20	W95187	Bovine genome-deri
20	172	100.0	98	21	B10346	Bovine oxytocin se
21	172	100.0	98	21	B10353	Bovine oxytocin se
22	163	94.8	31	18	W31384	Rat type G protein
23	163	94.8	31	20	W97233	Rat type ligand po
24	163	94.8	31	20	W87614	Rat 19P2 ligand.
25	163	94.8	31	20	W95173	Murine pituitary-d
26	163	94.8	31	20	W95174	Murine pituitary-d
27	163	94.8	31	21	B10355	Rat oxytocin secre
28	163	94.8	31	21	Y87504	Rat prolactin-rele
29	163	94.8	31	21	Y49292	19P2 ligand peptid
30	163	94.8	32	18	W31385	Rat type G protein
31	163	94.8	32	21	B10356	Rat oxytocin secre
32	163	94.8	33	18	W31386	Rat type G protein
33	163	94.8	33	21	B10357	Rat oxytocin secre
34	163	94.8	82	20	W95172	Murine pituitary-d
35	163	94.8	83	18	W31383	Rat type G protein
36	163	94.8	83	20	W97225	Rat type ligand po
37	163	94.8	83	21	B10354	Rat oxytocin secre
38	161	93.6	29	18	W31369	Bovine G protein-c
39	161	93.6	29	20	W95184	Bovine pituitary-d
40	158	91.9	31	18	W31391	Human type G prote
41	158	91.9	31	20	W97235	Human type ligand
42	158	91.9	31	20	W87615	Human 19P2 ligand.
43	158	91.9	31	21	B10362	Human oxytocin sec
44	158	91.9	31	21	Y49291	19P2 ligand peptid
45	158	91.9	32	18	W31392	Human type G prote

ALIGNMENTS

RESULT 1
W31371
ID W31371 standard; Peptide; 31 AA.
XX
AC W31371;
XX
DT 06-APR-1998 (first entry)
XX
DE Bovine G protein-coupled receptor ligand peptide fragment 1.
XX
KW G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
OS Bos taurus.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
DR N-PSDB; V02394.

XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2; Page 160; 258pp; English.
XX
CC This sequence represents a peptide fragment of a novel bovine pituitary
CC derived ligand corresponding to amino acid residues 23 to 53 of the
CC sequence in W31368 and is used in an assay to monitor ligand binding to
CC the G protein-coupled receptor protein. Pharmaceutical compositions
CC containing this ligand may be used as a pituitary function modulator, a
CC central nervous system modulator or a pancreatic function modulator. This
CC ligand could have specific applications as a prophylactic or therapeutic
CC agent for dementia, depression, hyperkinetic syndrome, disturbance of
CC consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
CC secretory disease, hyper- and polyphagia, hypercholesterolaemia,
CC hyperglyceridaemia, hyperlipidaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia,
CC amylotrophic lateral sclerosis, acute myocardial infarction,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
CC can also be developed to screen compounds which are capable of altering
CC the binding activity of the ligand thus affecting activation of the G
CC protein-coupled receptor protein.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 172; DB 18; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.9e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmeirtpdinpawyagrgirpvgrf 31

RESULT 2
W97218
ID W97218 standard; peptide; 31 AA.
XX
AC W97218;
XX
DT 06-MAY-1999 (first entry)
XX
DE Bovine pituitary-derived ligand polypeptide fragment.
XX
KW Bovine pituitary-derived ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypoovarianism; gonecyst cacogenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmeniopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castilo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.
XX
OS Bos sp.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI; 1999-105614/09.
XX
PT Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Claim 3; Page 135; 241pp; English.
XX
CC The present sequence represents a bovine pituitary-derived ligand
CC fragment. It is used in the course of the invention. The specification
CC describes an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypoovarianism, gonecyst cacogenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can by used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmeniopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castilo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, irruption mole, abortion, unthrifty fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 172; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.9e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmeirtpdinpawyagrgirpvgrf 31

RESULT 3
W87613
ID W87613 standard; Peptide; 31 AA.
XX
AC W87613;
XX
DT 29-MAR-1999 (first entry)
XX
DE Bovine 19P2 ligand.
XX
KW 19P2 ligand; G protein coupled receptor; pituitary;
KW prolactin releasing peptide; cattle; dementia; breast cancer;
KW therapy.
XX
OS Bos taurus.
XX
PN EP887417-A2.
XX
PD 30-DEC-1998.
XX
PF 25-JUN-1998; 98EP-0111725.
XX
PR 27-JUN-1997; 97JP-0172118.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Moriya T, Nishimura O, Suenaga M, Tanaka Y;
XX
DR WPI; 1999-047884/05.
XX
PT Producing a 19P2 pituitary G protein receptor ligand - by cleavage
PT of a fusion protein, useful for preventing and treating dementia,
PT breast cancer, renal failure and autoimmune disease
XX
PS Claim 5; Page 34; 56pp; English.

```
XX
CC   This is the amino acid sequence of the bovine pituitary G
CC   protein-coupled receptor ligand 19P2L.  A method suitable for
CC   commercial high-level production of 19P2L comprises expressing
CC   the ligand in host cells as a recombinant fusion protein e.g. with
CC   human basic fibroblast growth factor (see V83792-93) that has
CC   been modified to include an N-terminal cysteine residue.  The
CC   ligand is released from the fusion by cyanylation followed by
CC   ammonolysis.  19P2L has prolactin secretion-stimulating and (at
CC   high doses) prolactin secretion-inhibiting properties.  It can be
CC   used in the treatment and prevention of various diseases including:
CC   senile dementia, cerebrovascular dementia, and dementia associated
CC   with: genealogical disorders (e.g. Alzheimer's disease, Parkinson's
CC   disease, Pick's disease, Huntington's disease), infectious diseases
CC   (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or
CC   toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism,
CC   intoxication by drugs, metal and organic compounds), tumourigenic
CC   diseases (e.g. brain tumour), traumatic diseases (e.g. chronic
CC   subarachnoidal heamorrhage, and other types of dementia, depression,
CC   hyperactive child syndrome (microencephalopathy) and disturbance of
CC   consciousness.  It is also useful for prevention and treatment of
CC   diseases associated with prolactin hypo and hypersecretion
CC   respectively, including: hyperprolactinaemia, pituitary adenoma,
CC   breast cancer, infertility, impotence and autoimmune disease
CC   (hypersecretion disorders), and seminal vesicle hypoplasia,
CC   osteoporosis, menopausal syndrome and renal failure (hyposecretion
CC   disorders).  The 19P2 polypeptide/amide is also useful as a test
CC   reagent for study of the prolactin secretory function or as a
CC   lactogogue in mammalian farm animals.
XX
SQ   Sequence   31 AA;

  Query Match             100.0%;  Score 172;  DB 20;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
           |||||||||||||||||||||||||||||||
Db       1 srahqhsmeirtpdinpawyagrgirpvgrf 31


RESULT   4
W95188
ID   W95188 standard; peptide; 31 AA.
XX
AC   W95188;
XX
DT   10-MAR-1999  (first entry)
XX
DE   Bovine pituitary-derived ligand polypeptide fragment.
XX
KW   Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW   GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW   tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW   Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW   secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW   gene therapy; transgenic animal; bovine.
XX
OS   Bos sp.
XX
PN   WO9849295-A1.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-JP01923.
XX
PR   28-APR-1997;   97JP-0109974.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fukusumi S,  Hinuma S;
XX
DR   WPI; 1999-009423/01.
XX
PT   New polypeptide ligand for orphan G protein coupled receptors - used
PT   for treating disorders of central nervous system, pituitary and
PT   pancreas, and for drug screening
XX
PS   Example 19; Page 150; 206pp; English.
XX
CC   The invention relates to a murine pituitary-derived ligand polypeptide
CC   which is a ligand for the G-protein coupled orphan receptor designated
CC   GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC   the ligand polypeptide encoding DNA are used to produce a recombinant
CC   ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC   function of the pituitary, central nervous system, pancreas and other
CC   tissues and can be used to screen for agents that modulate binding of the
CC   polypeptide to the receptor; to quantify the amount of receptor in a
CC   sample and to raise antibodies. They may also be used therapeutically,
CC   e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC   diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC   diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC   rheumatoid arthritis, epilepsy and many others, also to improve post-
CC   operative nutritional status and as vasopressor. Transgenic animals
CC   carrying the ligand polypeptide encoding DNA or its mutein are used to
CC   study the function of the polypeptide-expressing genes, as models of
CC   disease, for drug screening and as source of cell lines. The ligand
CC   polypeptide DNA is used as a source of probes and primers; to identify
CC   related sequences; in receptor-binding assays; for production of Ab and
CC   antisera; in drug development; for gene therapy and to develop transgenic
CC   animals. The present sequence represents a bovine genome-derived ligand
CC   polypeptide fragment which is similar to the murine ligand-polypeptide.
XX
SQ   Sequence   31 AA;

  Query Match             100.0%;  Score 172;  DB 20;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
           |||||||||||||||||||||||||||||||
Db       1 srahqhsmeirtpdinpawyagrgirpvgrf 31


RESULT   5
B10347
ID   B10347 standard; peptide; 31 AA.
XX
AC   B10347;
XX
DT   24-NOV-2000  (first entry)
XX
DE   Bovine oxytocin secretion promoting peptide SEQ ID NO: 3.
XX
KW   Bovine; oxytocin secretion promoter; G protein-coupled receptor protein;
KW   treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW   caesarean section; artificial fertilization; galactostasis; goat; pig;
KW   veterinary medicine; milk production.
XX
OS   Bos taurus.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
```

```
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine  -
XX
PS   Claim 3; Page 50; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a bovine peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   31 AA;


  Query Match             100.0%;  Score 172;  DB 21;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmeirtpdinpawyagrgirpvgrf 31


RESULT   6
Y49290
ID   Y49290 standard; peptide; 31 AA.
XX
AC   Y49290;
XX
DT   22-FEB-2000  (first entry)
XX
DE   19P2 ligand peptide fragment.
XX
KW   Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW   pituitary; regulatory mechanism; central nervous system; pancreatic.
XX
OS   Bos sp.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   31
FT                   /note= "C-terminal amide"
XX
PN   WO9960112-A1.
XX
PD   25-NOV-1999.
XX
PF   20-MAY-1999;   99WO-JP02650.
XX
PR   21-MAY-1998;   98JP-0140293.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-039381/03.
XX
PT   New monoclonal antibodies,  useful in diagnosis, as drugs and  in
PT   studying diseases related to ligand abnormality  -
XX
PS   Disclosure; Page 26; 73pp; Japanese.
XX
CC   The invention provides a monoclonal antibody which has a specific
CC   reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC   derivative. The antibodies can be used in diagnosis or to treat or
CC   prevent diseases associated with abnormality in the pituitary function
CC   regulatory mechanism (e.g. promotion of prolactin secretion), central
CC   nervous regulatory mechanism, and pancreatic function regulatory
CC   mechanism. The antibody-based immunoassay can also be applied in
CC   clarifying the physiological functions of the ligand and its derivative.
CC   Sequences Y49290-302 represent peptide fragments of the 19P2 ligand.
XX
SQ   Sequence   31 AA;


  Query Match             100.0%;  Score 172;  DB 21;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmeirtpdinpawyagrgirpvgrf 31


RESULT   7
Y49298
ID   Y49298 standard; peptide; 31 AA.
XX
AC   Y49298;
XX
DT   22-FEB-2000  (first entry)
XX
DE   19P2 ligand peptide fragment.
XX
KW   Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW   pituitary; regulatory mechanism; central nervous system; pancreatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   31
FT                   /note= "C-terminal amide"
XX
PN   WO9960112-A1.
XX
PD   25-NOV-1999.
XX
PF   20-MAY-1999;   99WO-JP02650.
XX
PR   21-MAY-1998;   98JP-0140293.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-039381/03.
XX
PT   New monoclonal antibodies,  useful in diagnosis, as drugs and  in
PT   studying diseases related to ligand abnormality  -
XX
PS   Disclosure; Page 27; 73pp; Japanese.
XX
CC   The invention provides a monoclonal antibody which has a specific
CC   reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC   derivative. The antibodies can be used in diagnosis or to treat or
CC   prevent diseases associated with abnormality in the pituitary function
CC   regulatory mechanism (e.g. promotion of prolactin secretion), central
CC   nervous regulatory mechanism, and pancreatic function regulatory
CC   mechanism. The antibody-based immunoassay can also be applied in
CC   clarifying the physiological functions of the ligand and its derivative.
CC   Sequences Y49290-302 represent peptide fragments of the 19P2 ligand.
XX
SQ   Sequence   31 AA;


  Query Match             100.0%;  Score 172;  DB 21;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          ||||||||||||||||||||||||||||||||
Db      1 srahqhsmeirtpdinpawyagrgirpvgrf 31
```

```
RESULT   8
W31372
ID   W31372 standard; Peptide; 32 AA.
XX
AC   W31372;
XX
DT   06-APR-1998  (first entry)
XX
DE   Bovine G protein-coupled receptor ligand peptide fragment 2.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Bos taurus.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02395.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 160-161; 258pp; English.
XX
CC   This sequence represents a peptide fragment of a novel bovine pituitary
CC   derived ligand corresponding to amino acid residues 23 to 54 of the
CC   sequence in W31368 and is used in an assay to monitor ligand binding to
CC   the G protein-coupled receptor protein. Pharmaceutical compositions
CC   containing this ligand may be used as a pituitary function modulator, a
CC   central nervous system modulator or a pancreatic function modulator. This
CC   ligand could have specific applications as a prophylactic or therapeutic
CC   agent for dementia, depression, hyperkinetic syndrome, disturbance of
CC   consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
CC   secretory disease, hyper- and polyphagia, hypercholesterolaemia,
CC   hyperglyceridaemia, hyperlipidaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia,
CC   amylotrophic lateral sclerosis, acute myocardial infarction,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
CC   can also be developed to screen compounds which are capable of altering
CC   the binding activity of the ligand thus affecting activation of the G
CC   protein-coupled receptor protein.
XX
SQ   Sequence   32 AA;
```

```
  Query Match           100.0%;  Score 172;  DB 18;  Length 32;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          ||||||||||||||||||||||||||||||||
Db      1 srahqhsmeirtpdinpawyagrgirpvgrf 31
```

```
RESULT   9
W95189
ID   W95189 standard; peptide; 32 AA.
XX
AC   W95189;
XX
DT   10-MAR-1999  (first entry)
XX
DE   Bovine pituitary-derived ligand polypeptide fragment.
XX
KW   Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW   GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW   tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW   Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW   secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW   gene therapy; transgenic animal; bovine.
XX
OS   Bos sp.
XX
PN   WO9849295-A1.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-JP01923.
XX
PR   28-APR-1997;   97JP-0109974.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fukusumi S,  Hinuma S;
XX
DR   WPI; 1999-009423/01.
XX
PT   New polypeptide ligand for orphan G protein coupled receptors - used
PT   for treating disorders of central nervous system, pituitary and
PT   pancreas, and for drug screening
XX
PS   Example 19; Page 150; 206pp; English.
XX
CC   The invention relates to a murine pituitary-derived ligand polypeptide
CC   which is a ligand for the G-protein coupled orphan receptor designated
CC   GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC   the ligand polypeptide encoding DNA are used to produce a recombinant
CC   ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC   function of the pituitary, central nervous system, pancreas and other
CC   tissues and can be used to screen for agents that modulate binding of the
CC   polypeptide to the receptor; to quantify the amount of receptor in a
CC   sample and to raise antibodies. They may also be used therapeutically,
CC   e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC   diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC   diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC   rheumatoid arthritis, epilepsy and many others, also to improve post-
CC   operative nutritional status and as vasopressor. Transgenic animals
CC   carrying the ligand polypeptide encoding DNA or its mutein are used to
CC   study the function of the polypeptide-expressing genes, as models of
CC   disease, for drug screening and as source of cell lines. The ligand
CC   polypeptide DNA is used as a source of probes and primers; to identify
CC   related sequences; in receptor-binding assays; for production of Ab and
CC   antisera; in drug development; for gene therapy and to develop transgenic
CC   animals. The present sequence represents a bovine genome-derived ligand
CC   polypeptide fragment which is similar to the murine ligand-polypeptide.
XX
SQ   Sequence   32 AA;
```

```
  Query Match           100.0%;  Score 172;  DB 20;  Length 32;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-19;
```

```
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmeirtpdinpawyagrgirpvgrf 31


RESULT  10
B10348
ID   B10348 standard; peptide; 32 AA.
XX
AC   B10348;
XX
DT   24-NOV-2000  (first entry)
XX
DE   Bovine oxytocin secretion promoting peptide SEQ ID NO: 4.
XX
KW   Bovine; oxytocin secretion promoter; G protein-coupled receptor protein;
KW   treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW   caesarean section; artificial fertilization; galactostasis; goat; pig;
KW   veterinary medicine; milk production.
XX
OS   Bos taurus.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine -
XX
PS   Disclosure; Page 51; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a bovine peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   32 AA;


  Query Match            100.0%;  Score 172;  DB 21;  Length 32;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmeirtpdinpawyagrgirpvgrf 31


RESULT  11
W31373
ID   W31373 standard; Peptide; 33 AA.
XX
AC   W31373;
XX
DT   06-APR-1998  (first entry)
XX
DE   Bovine G protein-coupled receptor ligand peptide fragment 3.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Bos taurus.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02396.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 161; 258pp; English.
XX
CC   This sequence represents a peptide fragment of a novel bovine pituitary
CC   derived ligand corresponding to amino acid residues 23 to 55 of the
CC   sequence in W31368 and is used in an assay to monitor ligand binding to
CC   the G protein-coupled receptor protein. Pharmaceutical compositions
CC   containing this ligand may be used as a pituitary function modulator, a
CC   central nervous system modulator or a pancreatic function modulator. This
CC   ligand could have specific applications as a prophylactic or therapeutic
CC   agent for dementia, depression, hyperkinetic syndrome, disturbance of
CC   consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
CC   secretory disease, hyper- and polyphagia, hypercholesterolaemia,
CC   hyperglyceridaemia, hyperlipidaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia,
CC   amylotrophic lateral sclerosis, acute myocardial infarction,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
CC   can also be developed to screen compounds which are capable of altering
CC   the binding activity of the ligand thus affecting activation of the G
CC   protein-coupled receptor protein.
XX
SQ   Sequence   33 AA;


  Query Match            100.0%;  Score 172;  DB 18;  Length 33;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmeirtpdinpawyagrgirpvgrf 31


RESULT  12
W95190
ID   W95190 standard; peptide; 33 AA.
XX
AC   W95190;
XX
```

DT 10-MAR-1999 (first entry)
XX
DE Bovine pituitary-derived ligand polypeptide fragment.
XX
KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal; bovine.
XX
OS Bos sp.
XX
PN WO9849295-A1.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-JP01923.
XX
PR 28-APR-1997; 97JP-0109974.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fukusumi S, Hinuma S;
XX
DR WPI; 1999-009423/01.
XX
PT New polypeptide ligand for orphan G protein coupled receptors - used
PT for treating disorders of central nervous system, pituitary and
PT pancreas, and for drug screening
XX
PS Example 19; Page 150; 206pp; English.
XX
CC The invention relates to a murine pituitary-derived ligand polypeptide
CC which is a ligand for the G-protein coupled orphan receptor designated
CC GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC the ligand polypeptide encoding DNA are used to produce a recombinant
CC ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC function of the pituitary, central nervous system, pancreas and other
CC tissues and can be used to screen for agents that modulate binding of the
CC polypeptide to the receptor; to quantify the amount of receptor in a
CC sample and to raise antibodies. They may also be used therapeutically,
CC e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC rheumatoid arthritis, epilepsy and many others, also to improve post-
CC operative nutritional status and as vasopressor. Transgenic animals
CC carrying the ligand polypeptide encoding DNA or its mutein are used to
CC study the function of the polypeptide-expressing genes, as models of
CC disease, for drug screening and as source of cell lines. The ligand
CC polypeptide DNA is used as a source of probes and primers; to identify
CC related sequences; in receptor-binding assays; for production of Ab and
CC antisera; in drug development; for gene therapy and to develop transgenic
CC animals. The present sequence represents a bovine genome-derived ligand
CC polypeptide fragment which is similar to the murine ligand-polypeptide.
XX
SQ Sequence 33 AA;

Query Match 100.0%; Score 172; DB 20; Length 33;
Best Local Similarity 100.0%; Pred. No. 4.2e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmeirtpdinpawyagrgirpvgrf 31

RESULT 13
B10349
ID B10349 standard; peptide; 33 AA.
XX
AC B10349;
XX
DT 24-NOV-2000 (first entry)
XX
DE Bovine oxytocin secretion promoting peptide SEQ ID NO: 5.
XX
KW Bovine; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX
OS Bos taurus.
XX
PN WO200038704-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-JP07199.
XX
PR 25-DEC-1998; 98JP-0369585.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-452298/39.
XX
PT Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine -
XX
PS Disclosure; Page 51; 72pp; Japanese.
XX
CC This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a bovine peptide which acts as an oxytocin secretion
CC promoter.
XX
SQ Sequence 33 AA;

Query Match 100.0%; Score 172; DB 21; Length 33;
Best Local Similarity 100.0%; Pred. No. 4.2e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmeirtpdinpawyagrgirpvgrf 31

RESULT 14
Y49297
ID Y49297 standard; peptide; 33 AA.
XX
AC Y49297;
XX
DT 22-FEB-2000 (first entry)
XX
DE 19P2 ligand peptide fragment.
XX
KW Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW pituitary; regulatory mechanism; central nervous system; pancreatic.
XX
OS Bos sp.
XX
PN WO9960112-A1.

XX
PD 25-NOV-1999.
XX
PF 20-MAY-1999; 99WO-JP02650.
XX
PR 21-MAY-1998; 98JP-0140293.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-039381/03.
XX
PT New monoclonal antibodies, useful in diagnosis, as drugs and in
PT studying diseases related to ligand abnormality -
XX
PS Disclosure; Page 27; 73pp; Japanese.
XX
CC The invention provides a monoclonal antibody which has a specific
CC reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC derivative. The antibodies can be used in diagnosis or to treat or
CC prevent diseases associated with abnormality in the pituitary function
CC regulatory mechanism (e.g. promotion of prolactin secretion), central
CC nervous regulatory mechanism, and pancreatic function regulatory
CC mechanism. The antibody-based immunoassay can also be applied in
CC clarifying the physiological functions of the ligand and its derivative.
CC Sequences Y49290-302 represent peptide fragments of the 19P2 ligand.
XX
SQ Sequence 33 AA;

Query Match 100.0%; Score 172; DB 21; Length 33;
Best Local Similarity 100.0%; Pred. No. 4.2e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmeirtpdinpawyagrgirpvgrf 31

RESULT 15
W31382
ID W31382 standard; Protein; 98 AA.
XX
AC W31382;
XX
DT 06-APR-1998 (first entry)
XX
DE Bovine genome derived G protein-coupled receptor ligand.
XX
KW G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
OS Bos taurus.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 3; Page 177; 258pp; English.
XX
CC This sequence represents a novel bovine genome-derived ligand polypeptide
CC which is used in an assay to monitor ligand binding to the G
CC protein-coupled receptor protein. Pharmaceutical compositions containing
CC this ligand may be used as a pituitary function modulator, a central
CC nervous system modulator or a pancreatic function modulator. This ligand
CC could have specific applications as a prophylactic or therapeutic agent
CC for dementia, depression, hyperkinetic syndrome, disturbance of
CC consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
CC secretory disease, hyper- and polyphagia, hypercholesterolaemia,
CC hyperglyceridaemia, hyperlipidaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia,
CC amylotrophic lateral sclerosis, acute myocardial infarction,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
CC can also be developed to screen compounds which are capable of altering
CC the binding activity of the ligand thus affecting activation of the G
CC protein-coupled receptor protein.
XX
SQ Sequence 98 AA;

Query Match 100.0%; Score 172; DB 18; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.6e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 23 srahqhsmeirtpdinpawyagrgirpvgrf 53

Search completed: April 17, 2001, 15:38:41
Job time: 348 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:35:52 ; Search time 70.08 Seconds
(without alignments)
30.400 Million cell updates/sec

Title: US-09-446-543A-5
Perfect score: 172
Sequence: 1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	55	32.0	1236	2	T50904	Mg protoporphyrin
2	54.5	31.7	503	2	A82193	Sun/nucleolar prot
3	53.5	31.1	176	2	S67150	hypothetical prote
4	53	30.8	376	2	C75580	adenine deaminase-
5	52	30.2	1292	2	T31462	probable magnesium
6	51.5	29.9	664	2	F83376	conserved hypothet
7	51	29.7	294	2	T21075	hypothetical prote
8	51	29.7	430	1	B69009	conserved hypothet
9	51	29.7	798	2	S11210	probable unr prote
10	51	29.7	962	2	H69157	excinuclease ABC c
11	50.5	29.4	790	2	T47959	hypothetical prote
12	50	29.1	527	2	T33175	hypothetical prote
13	49	28.5	128	2	S76955	hypothetical prote
14	48.5	28.2	254	2	S76814	hypothetical prote
15	48.5	28.2	548	2	T47548	hypothetical prote
16	48.5	28.2	1882	2	S73484	hypothetical prote
17	48	27.9	220	2	C83292	probable glutathio
18	48	27.9	314	2	B70569	hypothetical prote
19	48	27.9	348	2	T21648	hypothetical prote
20	48	27.9	365	2	T39098	hypothetical prote
21	48	27.9	424	2	B38176	samB protein - Sal
22	48	27.9	772	2	T07958	protoporphyrin IX
23	48	27.9	960	2	A71315	excinuclease ABC c
24	48	27.9	1331	2	S75000	protoporphyrin IX
25	48	27.9	1379	2	S37310	protoporphyrin IX
26	48	27.9	1380	2	S64721	protoporphyrin IX
27	48	27.9	1381	2	S71288	protoporphyrin IX
28	48	27.9	1382	2	T01789	protoporphyrin IX
29	48	27.9	1383	2	T07126	magnesium chelatas
30	47	27.3	406	2	I46535	11-beta-hydroxyste
31	47	27.3	455	2	D70885	probable aldC prot
32	47	27.3	785	2	F69099	sensory transducti
33	47	27.3	940	2	A82329	excinuclease ABC,
34	47	27.3	972	2	A70619	excinuclease ABC c
35	47	27.3	1084	2	T33759	hypothetical prote
36	46.5	27.0	345	2	D84012	N-acetylglutamate
37	46.5	27.0	531	2	T35226	nitrate reductase
38	46.5	27.0	770	1	A38230	inorganic pyrophos
39	46.5	27.0	957	2	A84089	hypothetical prote
40	46	26.7	333	2	H82852	hydroxybenzoate oc
41	46	26.7	342	2	B64395	malic acid transpo
42	46	26.7	347	2	H64371	malic acid transpo
43	46	26.7	510	2	A64865	ycgB protein - Esc
44	46	26.7	798	2	S29815	N-ras upstream pro
45	46	26.7	836	2	A54269	protein-glutamine

ALIGNMENTS

RESULT 1
T50904
Mg protoporphyrin methyl transferase [imported] - Rubrivivax gelatinosus
C;Species: Rubrivivax gelatinosus
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 02-Sep-2000
C;Accession: T50904
R;Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.
submitted to the EMBL Data Library, November 1999
A;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photos
A;Reference number: Z25270
A;Accession: T50904
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1236 <NAG>
A;Cross-references: EMBL:AB034704; PIDN:BAA94057.1
A;Experimental source: strain IL144
C;Genetics:
A;Gene: bchH
C;Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase

Query Match 32.0%; Score 55; DB 2; Length 1236;
Best Local Similarity 34.4%; Pred. No. 7.9;
Matches 11; Conservative 6; Mismatches 9; Indels 6; Gaps 1;

```
Qy       3 AHQHSMEIRTPDINPAWYAG------RGIRPV 28
           : | ::| ||  :|| || |       |:| :
Db    1112 SEQVALETRTRMLNPKWYEGMLEHGYEGVRQI 1143
```

RESULT 2
A82193
Sun/nucleolar protein family protein VC1502 [imported] - Vibrio cholerae (group O1 st
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C;Accession: A82193
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: A82193
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-503 <HEI>
A;Cross-references: GB:AE004228; GB:AE003852; NID:g9655997; PIDN:AAF94657.1; GSPDB:GN
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC1502
A;Map position: 1

```
  Query Match             31.7%;  Score 54.5;  DB 2;  Length 503;
  Best Local Similarity   36.7%;  Pred. No. 3.5;
  Matches   11;  Conservative    6;  Mismatches   12;  Indels    1;  Gaps    1;

Qy       1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGR 30
           | :  ||:|: |      |: || :|| |:
Db     435 SSSASHSVELDTTQAR-EWFMGRDVRPEGQ 463


RESULT   3
S67150
hypothetical protein YOR253w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein O5315
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 05-Nov-1999
C;Accession: S67150
R;Jauniaux, J.C.; Poirey, R.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67143
A;Accession: S67150
A;Molecule type: DNA
A;Residues: 1-176 <JAU>
A;Cross-references: EMBL:Z75161; NID:g1420572; PID:e252411; PID:g1420573; GSPDB:GN00015;
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YOR253w
A;Map position: 15R


  Query Match             31.1%;  Score 53.5;  DB 2;  Length 176;
  Best Local Similarity   31.2%;  Pred. No. 1.6;
  Matches   10;  Conservative    5;  Mismatches   14;  Indels    3;  Gaps    1;

Qy       1 SRAHQHSMEIRTP---DINPAWYAGRGIRPVG 29
           |  |||:: :  |   |:   |:   |   ||
Db     120 SECHQHNVFVYLPAVDDLTKQWFIAHGFEQVG 151


RESULT   4
C75580
adenine deaminase-related protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: C75580
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: C75580
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-376 <WHI>
A;Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12376.1; PID:g646067
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0268
A;Map position: 2


  Query Match             30.8%;  Score 53;  DB 2;  Length 376;
  Best Local Similarity   40.0%;  Pred. No. 4.3;
  Matches   12;  Conservative    6;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGR 30
           :||  |: ::  || : |  ||:  || ||
Db     177 ARALAHAPDVSRPDRHRACGAGQRRRPAGR 206


RESULT   5
T31462
```

```
probable magnesium chelatase (EC 4.99.1.-) chain H BchH - Heliobacillus mobilis
C;Species: Heliobacillus mobilis
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 15-Sep-2000
C;Accession: T31462
R;Xiong, J.; Inoue, K.; Bauer, C.E.
Proc. Natl. Acad. Sci. U.S.A. 95, 14851-14856, 1998
A;Title: Tracking molecular evolution of photosynthesis by characterization of a majo
A;Reference number: Z21036; MUID:99061957
A;Accession: T31462
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1292 <XIO>
A;Cross-references: EMBL:AF080002; NID:g3820536; PID:g3820560; PIDN:AAC84033.1
C;Genetics:
A;Gene: bchH
C;Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase
C;Keywords: lyase


  Query Match             30.2%;  Score 52;  DB 2;  Length 1292;
  Best Local Similarity   38.5%;  Pred. No. 23;
  Matches   10;  Conservative    3;  Mismatches    7;  Indels    6;  Gaps    1;

Qy       9 EIRTPDINPAWYAG------RGIRPV 28
           | ||  :|| || |       |:| :
Db    1176 ETRTKTLNPKWYEGMLKHGYEGVREI 1201


RESULT   6
F83376
conserved hypothetical protein PA2151 [imported] - Pseudomonas aeruginosa (strain PAO
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83376
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: F83376
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-664 <STO>
A;Cross-references: GB:AE004642; GB:AE004091; NID:g9948163; PIDN:AAG05539.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2151


  Query Match             29.9%;  Score 51.5;  DB 2;  Length 664;
  Best Local Similarity   55.6%;  Pred. No. 13;
  Matches   10;  Conservative    2;  Mismatches    5;  Indels    1;  Gaps    1;

Qy      10 IRTPDINPAWYAGRGIRP 27
           : |||||| |:  |  ||
Db     476 VNTPDINP-WFLQRSGRP 492


RESULT   7
T21075
hypothetical protein F19H6.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T21075; T21124
R;McMurray, A.
submitted to the EMBL Data Library, August 1995
A;Reference number: Z19368
A;Accession: T21075
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-294 <WIL>
A;Cross-references: EMBL:Z50873; PIDN:CAA90762.1; GSPDB:GN00028; CESP:F19H6.1
```

```
A;Experimental source: clone F17E5
R;McMurray, A.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19378
A;Accession: T21124
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-294 <WI2>
A;Cross-references: EMBL:Z68115; PIDN:CAA92169.1; GSPDB:GN00028; CESP:F19H6.1
A;Experimental source: clone F19H6
C;Genetics:
A;Gene: CESP:F19H6.1
A;Map position: X
A;Introns: 55/3; 85/2; 113/3; 150/1; 192/3; 266/2
C;Superfamily: kinase-related transforming protein; protein kinase homology


  Query Match             29.7%;  Score 51;  DB 2;  Length 294;
  Best Local Similarity   44.8%;  Pred. No. 6.5;
  Matches   13;  Conservative    2;  Mismatches   12;  Indels    2;  Gaps    1;

Qy       3 AHQHSMEIRTPDINPA--WYAGRGIRPVG 29
           || ||  |   || ||  :  | ||  :|
Db     139 AHMHSKRIMHRDIKPANVFITGNGIVKLG 167


RESULT   8
B69009
conserved hypothetical protein MTH1070 - Methanobacterium thermoautotrophicum (strain De
C;Species: Methanobacterium thermoautotrophicum
C;Date: 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change 23-Jul-1999
C;Accession: B69009
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514
A;Accession: B69009
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-430 <MTH>
A;Cross-references: GB:AE000877; GB:AE000666; NID:g2622157; PIDN:AAB85559.1; PID:g262216
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1070
C;Superfamily: conserved hypothetical protein MTH1070


  Query Match             29.7%;  Score 51;  DB 1;  Length 430;
  Best Local Similarity   69.2%;  Pred. No. 9.8;
  Matches    9;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      14 DINPAWYAGRGIR 26
           |||| | |||  |
Db     191 DINPEWVAGRACR 203


RESULT   9
S11210
probable unr protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C;Accession: S11210
R;Jeffers, M.; Paciucci, R.; Pellicer, A.
Nucleic Acids Res. 18, 4891-4899, 1990
A;Title: Characterization of unr; a gene closely linked to N-ras.
A;Reference number: S11210; MUID:90370473
A;Accession: S11210
A;Molecule type: mRNA
A;Residues: 1-798 <JEF>
A;Cross-references: EMBL:X52311; NID:g57454; PIDN:CAA36549.1; PID:g57455
C;Keywords: DNA binding


  Query Match             29.7%;  Score 51;  DB 2;  Length 798;
  Best Local Similarity   43.5%;  Pred. No. 19;
  Matches   10;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       6 HSMEIRTPDINPAWYAGRGIRPV 28
           ||:   | : ||  |:|: |||:
Db     583 HSVNGITEEANPTIYSGKVIRPL 605


RESULT  10
H69157
excinuclease ABC chain A - Methanobacterium thermoautotrophicum (strain Delta H)
N;Alternate names: uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 12-Nov-1999
C;Accession: H69157
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: H69157
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-962 <MTH>
A;Cross-references: GB:AE000828; GB:AE000666; NID:g2621504; PIDN:AAB84949.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH443
A;Start codon: TTG
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; P-loop
F;38-45/Region: nucleotide-binding motif A (P-loop)
F;632-915/Domain: ATP-binding cassette homology <ABCE>
F;649-656/Region: nucleotide-binding motif A (P-loop)


  Query Match             29.7%;  Score 51;  DB 2;  Length 962;
  Best Local Similarity   39.5%;  Pred. No. 23;
  Matches   15;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYAG----------------RGIRPVGRF 31
           |||  ||| | |                || || |||
Db     703 RTPRSNPATYTGVFTHIRELFAQTPEARKRGYRP-GRF 739


RESULT  11
T47959
hypothetical protein F15G16.60 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47959
R;De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Q
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z24480
A;Accession: T47959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-790 <DEH>
A;Cross-references: EMBL:AL132959
A;Experimental source: cultivar Columbia; BAC clone F15G16
C;Genetics:
A;Map position: 3
A;Introns: 39/1; 678/2; 698/3; 773/2
A;Note: F15G16.60


  Query Match             29.4%;  Score 50.5;  DB 2;  Length 790;
  Best Local Similarity   39.3%;  Pred. No. 22;
```

Matches 11; Conservative 4; Mismatches 12; Indels 1; Gaps 1;

```
Qy      4 HQHSMEIRTPDINPAWYAGRGIRPVGRF 31
          |:  |    |  ||  |  ||::| ||:
Db    358 HESVMGFAPPH-NPRTYGSRGLQPHGRW 384
```

RESULT 12
T33175
hypothetical protein C24G6.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33175
R;Greco, T.; Bradshaw, H.; Keppler, D.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid C24G6.
A;Reference number: Z21298
A;Accession: T33175
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-527 <GRE>
A;Cross-references: EMBL:AF067936; PIDN:AAC19213.1; GSPDB:GN00023; CESP:C24G6.6
A;Experimental source: strain Bristol N2; clone C24G6
C;Genetics:
A;Gene: CESP:C24G6.6
A;Map position: 5
A;Introns: 20/3; 77/1; 129/2; 208/3; 470/2

Query Match 29.1%; Score 50; DB 2; Length 527;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy     13 PDINPAWYAGRG 24
          |::  |||||||
Db    370 PNVLSAWYAGRG 381
```

RESULT 13
S76955
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S76955
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S76955
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-128 <KAN>
A;Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BAA18867.1; PID:d101960
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 28.5%; Score 49; DB 2; Length 128;
Best Local Similarity 33.3%; Pred. No. 5.2;
Matches 11; Conservative 5; Mismatches 13; Indels 4; Gaps 1;

```
Qy      1 SRAHQHSMEIRTPDINPAWYAGR----GIRPVG 29
          |  :  |::  :|| :  ||  |    || | |
Db     11 SPENNQSLDCSSPDPSRKWYQARLFSFGITPTG 43
```

RESULT 14
S76814
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S76814
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S76814
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <KAN>
A;Cross-references: EMBL:D90916; GB:AB001339; NID:g1653715; PIDN:BAA18726.1; PID:g165
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: yeast SOL3 protein

Query Match 28.2%; Score 48.5; DB 2; Length 254;
Best Local Similarity 42.3%; Pred. No. 13;
Matches 11; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

```
Qy      5 QHSM-EIRTPDINPAWYAGRGIRPVG 29
          ||:: ||  |: :|  |  | |:| |
Db    213 QHALGEIFAPEADPQQYPARFIQPQG 238
```

RESULT 15
T47548
hypothetical protein F8J2.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47548
R;Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M
Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24458
A;Accession: T47548
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-548 <NYA>
A;Cross-references: EMBL:AL132969
A;Experimental source: cultivar Columbia; BAC clone F8J2
C;Genetics:
A;Map position: 3
A;Introns: 78/1; 143/1; 242/2; 377/3; 428/2; 447/3; 470/3; 502/3; 532/3
A;Note: F8J2.80

Query Match 28.2%; Score 48.5; DB 2; Length 548;
Best Local Similarity 35.5%; Pred. No. 30;
Matches 11; Conservative 3; Mismatches 10; Indels 7; Gaps 1;

```
Qy      1 SRAHQ-------HSMEIRTPDINPAWYAGRG 24
          |  ||       | ::   |   |:|| |||
Db    118 SLLHQSPMHFLHHPLQHSFPHHQPSWYWGRG 148
```

Search completed: April 17, 2001, 15:45:51
Job time: 599 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:38:41 ; Search time 116.94 Seconds
(without alignments)
9.777 Million cell updates/sec

Title: US-09-446-543A-8
Perfect score: 115
Sequence: 1 TPDINPAWYAGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	115	100.0	20	18	W31374	Bovine G protein-c
2	115	100.0	20	20	W97232	Bovine pituitary-d
3	115	100.0	20	20	W95191	Bovine pituitary-d
4	115	100.0	20	21	B10350	Bovine oxytocin se
5	115	100.0	20	21	Y49301	19P2 ligand peptid
6	115	100.0	21	18	W31375	Bovine G protein-c
7	115	100.0	21	20	W95192	Bovine pituitary-d
8	115	100.0	21	21	B10351	Bovine oxytocin se
9	115	100.0	22	18	W31376	Bovine G protein-c
10	115	100.0	22	20	W95193	Bovine pituitary-d
11	115	100.0	22	21	B10352	Bovine oxytocin se
12	115	100.0	31	18	W31371	Bovine G protein-c
13	115	100.0	31	20	W97218	Bovine pituitary-d
14	115	100.0	31	20	W87613	Bovine 19P2 ligand
15	115	100.0	31	20	W95188	Bovine pituitary-d
16	115	100.0	31	21	B10347	Bovine oxytocin se
17	115	100.0	31	21	Y49290	19P2 ligand peptid
18	115	100.0	31	21	Y49298	19P2 ligand peptid
19	115	100.0	32	18	W31372	Bovine G protein-c
20	115	100.0	32	20	W95189	Bovine pituitary-d
21	115	100.0	32	21	B10348	Bovine oxytocin se
22	115	100.0	33	18	W31373	Bovine G protein-c
23	115	100.0	33	20	W95190	Bovine pituitary-d
24	115	100.0	33	21	B10349	Bovine oxytocin se
25	115	100.0	33	21	Y49297	19P2 ligand peptid
26	115	100.0	98	18	W31382	Bovine genome deri
27	115	100.0	98	18	W31368	Bovine G protein-c
28	115	100.0	98	20	W97224	Bovine genome-driv
29	115	100.0	98	20	W97217	Bovine pituitary-d
30	115	100.0	98	20	W95187	Bovine genome-deri
31	115	100.0	98	21	B10346	Bovine oxytocin se
32	115	100.0	98	21	B10353	Bovine oxytocin se
33	111	96.5	20	18	W31387	Rat type G protein
34	111	96.5	20	20	W97234	Rat type ligand po
35	111	96.5	20	20	W95175	Murine pituitary-d
36	111	96.5	20	21	B10358	Rat oxytocin secre
37	111	96.5	20	21	Y49302	19P2 ligand peptid
38	111	96.5	21	18	W31388	Rat type G protein
39	111	96.5	21	21	B10359	Rat oxytocin secre
40	111	96.5	22	18	W31389	Rat type G protein
41	111	96.5	22	21	B10360	Rat oxytocin secre
42	111	96.5	31	18	W31384	Rat type G protein
43	111	96.5	31	20	W97233	Rat type ligand po
44	111	96.5	31	20	W87614	Rat 19P2 ligand.
45	111	96.5	31	21	Y87504	Rat prolactin-rele

ALIGNMENTS

RESULT 1
W31374
ID W31374 standard; Peptide; 20 AA.
XX
AC W31374;
XX
DT 06-APR-1998 (first entry)
XX
DE Bovine G protein-coupled receptor ligand peptide fragment 4.
XX
KW G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
OS Bos taurus.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
DR N-PSDB; V02397.

XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2; Page 161; 258pp; English.
XX
CC This sequence represents a peptide fragment of a novel bovine pituitary
CC derived ligand corresponding to amino acid residues 34 to 53 of the
CC sequence in W31368 and is used in an assay to monitor ligand binding to
CC the G protein-coupled receptor protein. Pharmaceutical compositions
CC containing this ligand may be used as a pituitary function modulator, a
CC central nervous system modulator or a pancreatic function modulator. This
CC ligand could have specific applications as a prophylactic or therapeutic
CC agent for dementia, depression, hyperkinetic syndrome, disturbance of
CC consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
CC secretory disease, hyper- and polyphagia, hypercholesterolaemia,
CC hyperglyceridaemia, hyperlipidaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia,
CC amylotrophic lateral sclerosis, acute myocardial infarction,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
CC can also be developed to screen compounds which are capable of altering
CC the binding activity of the ligand thus affecting activation of the G
CC protein-coupled receptor protein.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 115; DB 18; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawyagrgirpvgrf 20

RESULT 2
W97232
ID W97232 standard; peptide; 20 AA.
XX
AC W97232;
XX
DT 06-MAY-1999 (first entry)
XX
DE Bovine pituitary-derived ligand polypeptide fragment.
XX
KW Bovine pituitary-derived ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypoovarianism; gonecyst cacogenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmeniopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castilo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.
XX
OS Bos sp.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI; 1999-105614/09.
XX
PT Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Claim 3; Page 136; 241pp; English.
XX
CC The present sequence represents a bovine pituitary-derived ligand
CC fragment. It is used in the course of the invention. The specification
CC describes an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypoovarianism, gonecyst cacogenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can by used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmeniopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castilo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, irruption mole, abortion, unthrifty fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 115; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawyagrgirpvgrf 20

RESULT 3
W95191
ID W95191 standard; peptide; 20 AA.
XX
AC W95191;
XX
DT 10-MAR-1999 (first entry)
XX
DE Bovine pituitary-derived ligand polypeptide fragment.
XX
KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal; bovine.
XX
OS Bos sp.
XX
PN WO9849295-A1.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-JP01923.
XX
PR 28-APR-1997; 97JP-0109974.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fukusumi S, Hinuma S;
XX
DR WPI; 1999-009423/01.
XX
PT New polypeptide ligand for orphan G protein coupled receptors - used

PT for treating disorders of central nervous system, pituitary and
PT pancreas, and for drug screening
XX
PS Example 19; Page 151; 206pp; English.
XX
CC The invention relates to a murine pituitary-derived ligand polypeptide
CC which is a ligand for the G-protein coupled orphan receptor designated
CC GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC the ligand polypeptide encoding DNA are used to produce a recombinant
CC ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC function of the pituitary, central nervous system, pancreas and other
CC tissues and can be used to screen for agents that modulate binding of the
CC polypeptide to the receptor; to quantify the amount of receptor in a
CC sample and to raise antibodies. They may also be used therapeutically,
CC e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC rheumatoid arthritis, epilepsy and many others, also to improve post-
CC operative nutritional status and as vasopressor. Transgenic animals
CC carrying the ligand polypeptide encoding DNA or its mutein are used to
CC study the function of the polypeptide-expressing genes, as models of
CC disease, for drug screening and as source of cell lines. The ligand
CC polypeptide DNA is used as a source of probes and primers; to identify
CC related sequences; in receptor-binding assays; for production of Ab and
CC antisera; in drug development; for gene therapy and to develop transgenic
CC animals. The present sequence represents a bovine genome-derived ligand
CC polypeptide fragment which is similar to the murine ligand-polypeptide.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 115; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawyagrgirpvgrf 20

RESULT 4
B10350
ID B10350 standard; peptide; 20 AA.
XX
AC B10350;
XX
DT 24-NOV-2000 (first entry)
XX
DE Bovine oxytocin secretion promoting peptide SEQ ID NO: 6.
XX
KW Bovine; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX
OS Bos taurus.
XX
PN WO200038704-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-JP07199.
XX
PR 25-DEC-1998; 98JP-0369585.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-452298/39.
XX
PT Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine -
XX
PS Claim 5; Page 51; 72pp; Japanese.
XX
CC This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a bovine peptide which acts as an oxytocin secretion
CC promoter.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 115; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawyagrgirpvgrf 20

RESULT 5
Y49301
ID Y49301 standard; peptide; 20 AA.
XX
AC Y49301;
XX
DT 22-FEB-2000 (first entry)
XX
DE 19P2 ligand peptide fragment.
XX
KW Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW pituitary; regulatory mechanism; central nervous system; pancreatic.
XX
OS Bos sp.
XX
FH Key Location/Qualifiers
FT Modified-site 20
FT /note= "C-terminal amide"
XX
PN WO9960112-A1.
XX
PD 25-NOV-1999.
XX
PF 20-MAY-1999; 99WO-JP02650.
XX
PR 21-MAY-1998; 98JP-0140293.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-039381/03.
XX
PT New monoclonal antibodies, useful in diagnosis, as drugs and in
PT studying diseases related to ligand abnormality -
XX
PS Disclosure; Page 27; 73pp; Japanese.
XX
CC The invention provides a monoclonal antibody which has a specific
CC reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC derivative. The antibodies can be used in diagnosis or to treat or
CC prevent diseases associated with abnormality in the pituitary function
CC regulatory mechanism (e.g. promotion of prolactin secretion), central
CC nervous regulatory mechanism, and pancreatic function regulatory
CC mechanism. The antibody-based immunoassay can also be applied in

```
CC   clarifying the physiological functions of the ligand and its derivative.
CC   Sequences Y49290-302 represent peptide fragments of the 19P2 ligand.
XX
SQ   Sequence   20 AA;


  Query Match             100.0%;  Score 115;  DB 21;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawyagrgirpvgrf 20


RESULT   6
W31375
ID   W31375 standard; Peptide; 21 AA.
XX
AC   W31375;
XX
DT   06-APR-1998  (first entry)
XX
DE   Bovine G protein-coupled receptor ligand peptide fragment 5.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Bos taurus.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02398.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 162; 258pp; English.
XX
CC   This sequence represents a peptide fragment of a novel bovine pituitary
CC   derived ligand corresponding to amino acid residues 34 to 54 of the
CC   sequence in W31368 and is used in an assay to monitor ligand binding to
CC   the G protein-coupled receptor protein. Pharmaceutical compositions
CC   containing this ligand may be used as a pituitary function modulator, a
CC   central nervous system modulator or a pancreatic function modulator. This
CC   ligand could have specific applications as a prophylactic or therapeutic
CC   agent for dementia, depression, hyperkinetic syndrome, disturbance of
CC   consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
CC   secretory disease, hyper- and polyphagia, hypercholesterolaemia,
CC   hyperglyceridaemia, hyperlipidaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia,
CC   amylotrophic lateral sclerosis, acute myocardial infarction,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
CC   can also be developed to screen compounds which are capable of altering
CC   the binding activity of the ligand thus affecting activation of the G
CC   protein-coupled receptor protein.
XX
SQ   Sequence   21 AA;


  Query Match             100.0%;  Score 115;  DB 18;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawyagrgirpvgrf 20


RESULT   7
W95192
ID   W95192 standard; peptide; 21 AA.
XX
AC   W95192;
XX
DT   10-MAR-1999  (first entry)
XX
DE   Bovine pituitary-derived ligand polypeptide fragment.
XX
KW   Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW   GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW   tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW   Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW   secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW   gene therapy; transgenic animal; bovine.
XX
OS   Bos sp.
XX
PN   WO9849295-A1.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-JP01923.
XX
PR   28-APR-1997;   97JP-0109974.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fukusumi S,  Hinuma S;
XX
DR   WPI; 1999-009423/01.
XX
PT   New polypeptide ligand for orphan G protein coupled receptors - used
PT   for treating disorders of central nervous system, pituitary and
PT   pancreas, and for drug screening
XX
PS   Example 19; Page 151; 206pp; English.
XX
CC   The invention relates to a murine pituitary-derived ligand polypeptide
CC   which is a ligand for the G-protein coupled orphan receptor designated
CC   GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC   the ligand polypeptide encoding DNA are used to produce a recombinant
CC   ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC   function of the pituitary, central nervous system, pancreas and other
CC   tissues and can be used to screen for agents that modulate binding of the
CC   polypeptide to the receptor; to quantify the amount of receptor in a
CC   sample and to raise antibodies. They may also be used therapeutically,
CC   e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC   diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC   diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC   rheumatoid arthritis, epilepsy and many others, also to improve post-
CC   operative nutritional status and as vasopressor. Transgenic animals
CC   carrying the ligand polypeptide encoding DNA or its mutein are used to
CC   study the function of the polypeptide-expressing genes, as models of
CC   disease, for drug screening and as source of cell lines. The ligand
CC   polypeptide DNA is used as a source of probes and primers; to identify
CC   related sequences; in receptor-binding assays; for production of Ab and
```

CC antisera; in drug development; for gene therapy and to develop transgenic
CC animals. The present sequence represents a bovine genome-derived ligand
CC polypeptide fragment which is similar to the murine ligand-polypeptide.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 115; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawyagrgirpvgrf 20

RESULT 8
B10351
ID B10351 standard; peptide; 21 AA.
XX
AC B10351;
XX
DT 24-NOV-2000 (first entry)
XX
DE Bovine oxytocin secretion promoting peptide SEQ ID NO: 7.
XX
KW Bovine; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX
OS Bos taurus.
XX
PN WO200038704-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-JP07199.
XX
PR 25-DEC-1998; 98JP-0369585.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-452298/39.
XX
PT Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine -
XX
PS Disclosure; Page 52; 72pp; Japanese.
XX
CC This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a bovine peptide which acts as an oxytocin secretion
CC promoter.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 115; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawyagrgirpvgrf 20

RESULT 9
W31376
ID W31376 standard; Peptide; 22 AA.
XX
AC W31376;
XX
DT 06-APR-1998 (first entry)
XX
DE Bovine G protein-coupled receptor ligand peptide fragment 6.
XX
KW G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
OS Bos taurus.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
DR N-PSDB; V02399.
XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2; Page 162; 258pp; English.
XX
CC This sequence represents a peptide fragment of a novel bovine pituitary
CC derived ligand corresponding to amino acid residues 34 to 55 of the
CC sequence in W31368 and is used in an assay to monitor ligand binding to
CC the G protein-coupled receptor protein. Pharmaceutical compositions
CC containing this ligand may be used as a pituitary function modulator, a
CC central nervous system modulator or a pancreatic function modulator. This
CC ligand could have specific applications as a prophylactic or therapeutic
CC agent for dementia, depression, hyperkinetic syndrome, disturbance of
CC consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
CC secretory disease, hyper- and polyphagia, hypercholesterolaemia,
CC hyperglyceridaemia, hyperlipidaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia,
CC amylotrophic lateral sclerosis, acute myocardial infarction,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
CC can also be developed to screen compounds which are capable of altering
CC the binding activity of the ligand thus affecting activation of the G
CC protein-coupled receptor protein.
XX
SQ Sequence 22 AA;

Query Match 100.0%; Score 115; DB 18; Length 22;
Best Local Similarity 100.0%; Pred. No. 8.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||

```
Db      1 tpdinpawyagrgirpvgrf 20


RESULT  10
W95193
ID   W95193 standard; peptide; 22 AA.
XX
AC   W95193;
XX
DT   10-MAR-1999  (first entry)
XX
DE   Bovine pituitary-derived ligand polypeptide fragment.
XX
KW   Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW   GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW   tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW   Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW   secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW   gene therapy; transgenic animal; bovine.
XX
OS   Bos sp.
XX
PN   WO9849295-A1.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-JP01923.
XX
PR   28-APR-1997;   97JP-0109974.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fukusumi S,  Hinuma S;
XX
DR   WPI; 1999-009423/01.
XX
PT   New polypeptide ligand for orphan G protein coupled receptors - used
PT   for treating disorders of central nervous system, pituitary and
PT   pancreas, and for drug screening
XX
PS   Example 19; Page 151; 206pp; English.
XX
CC   The invention relates to a murine pituitary-derived ligand polypeptide
CC   which is a ligand for the G-protein coupled orphan receptor designated
CC   GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC   the ligand polypeptide encoding DNA are used to produce a recombinant
CC   ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC   function of the pituitary, central nervous system, pancreas and other
CC   tissues and can be used to screen for agents that modulate binding of the
CC   polypeptide to the receptor; to quantify the amount of receptor in a
CC   sample and to raise antibodies. They may also be used therapeutically,
CC   e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC   diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC   diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC   rheumatoid arthritis, epilepsy and many others, also to improve post-
CC   operative nutritional status and as vasopressor. Transgenic animals
CC   carrying the ligand polypeptide encoding DNA or its mutein are used to
CC   study the function of the polypeptide-expressing genes, as models of
CC   disease, for drug screening and as source of cell lines. The ligand
CC   polypeptide DNA is used as a source of probes and primers; to identify
CC   related sequences; in receptor-binding assays; for production of Ab and
CC   antisera; in drug development; for gene therapy and to develop transgenic
CC   animals. The present sequence represents a bovine genome-derived ligand
CC   polypeptide fragment which is similar to the murine ligand-polypeptide.
XX
SQ   Sequence   22 AA;


  Query Match            100.0%;  Score 115;  DB 20;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 8.5e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawyagrgirpvgrf 20


RESULT  11
B10352
ID   B10352 standard; peptide; 22 AA.
XX
AC   B10352;
XX
DT   24-NOV-2000  (first entry)
XX
DE   Bovine oxytocin secretion promoting peptide SEQ ID NO: 8.
XX
KW   Bovine; oxytocin secretion promoter; G protein-coupled receptor protein;
KW   treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW   caesarean section; artificial fertilization; galactostasis; goat; pig;
KW   veterinary medicine; milk production.
XX
OS   Bos taurus.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine  -
XX
PS   Disclosure; Page 52; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a bovine peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   22 AA;


  Query Match            100.0%;  Score 115;  DB 21;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 8.5e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawyagrgirpvgrf 20


RESULT  12
W31371
ID   W31371 standard; Peptide; 31 AA.
XX
AC   W31371;
XX
DT   06-APR-1998  (first entry)
```

XX
DE Bovine G protein-coupled receptor ligand peptide fragment 1.
XX
KW G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
OS Bos taurus.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
DR N-PSDB; V02394.
XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2; Page 160; 258pp; English.
XX
CC This sequence represents a peptide fragment of a novel bovine pituitary
CC derived ligand corresponding to amino acid residues 23 to 53 of the
CC sequence in W31368 and is used in an assay to monitor ligand binding to
CC the G protein-coupled receptor protein. Pharmaceutical compositions
CC containing this ligand may be used as a pituitary function modulator, a
CC central nervous system modulator or a pancreatic function modulator. This
CC ligand could have specific applications as a prophylactic or therapeutic
CC agent for dementia, depression, hyperkinetic syndrome, disturbance of
CC consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
CC secretory disease, hyper- and polyphagia, hypercholesterolaemia,
CC hyperglyceridaemia, hyperlipidaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia,
CC amylotrophic lateral sclerosis, acute myocardial infarction,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
CC can also be developed to screen compounds which are capable of altering
CC the binding activity of the ligand thus affecting activation of the G
CC protein-coupled receptor protein.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 115; DB 18; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.2e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawyagrgirpvgrf 31

RESULT 13
W97218
ID W97218 standard; peptide; 31 AA.
XX
AC W97218;
XX
DT 06-MAY-1999 (first entry)
XX
DE Bovine pituitary-derived ligand polypeptide fragment.
XX
KW Bovine pituitary-derived ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypoovarianism; gonecyst cacogenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmeniopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castilo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.
XX
OS Bos sp.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI; 1999-105614/09.
XX
PT Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Claim 3; Page 135; 241pp; English.
XX
CC The present sequence represents a bovine pituitary-derived ligand
CC fragment. It is used in the course of the invention. The specification
CC describes an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypoovarianism, gonecyst cacogenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can by used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmeniopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castilo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, irruption mole, abortion, unthrifty fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 115; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.2e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawyagrgirpvgrf 31

RESULT 14
W87613
ID W87613 standard; Peptide; 31 AA.
XX
AC W87613;
XX
DT 29-MAR-1999 (first entry)
XX

DE Bovine 19P2 ligand.
XX
KW 19P2 ligand; G protein coupled receptor; pituitary;
KW prolactin releasing peptide; cattle; dementia; breast cancer;
KW therapy.
XX
OS Bos taurus.
XX
PN EP887417-A2.
XX
PD 30-DEC-1998.
XX
PF 25-JUN-1998; 98EP-0111725.
XX
PR 27-JUN-1997; 97JP-0172118.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Moriya T, Nishimura O, Suenaga M, Tanaka Y;
XX
DR WPI; 1999-047884/05.
XX
PT Producing a 19P2 pituitary G protein receptor ligand - by cleavage
PT of a fusion protein, useful for preventing and treating dementia,
PT breast cancer, renal failure and autoimmune disease
XX
PS Claim 5; Page 34; 56pp; English.
XX
CC This is the amino acid sequence of the bovine pituitary G
CC protein-coupled receptor ligand 19P2L. A method suitable for
CC commercial high-level production of 19P2L comprises expressing
CC the ligand in host cells as a recombinant fusion protein e.g. with
CC human basic fibroblast growth factor (see V83792-93) that has
CC been modified to include an N-terminal cysteine residue. The
CC ligand is released from the fusion by cyanylation followed by
CC ammonolysis. 19P2L has prolactin secretion-stimulating and (at
CC high doses) prolactin secretion-inhibiting properties. It can be
CC used in the treatment and prevention of various diseases including:
CC senile dementia, cerebrovascular dementia, and dementia associated
CC with: genealogical disorders (e.g. Alzheimer's disease, Parkinson's
CC disease, Pick's disease, Huntington's disease), infectious diseases
CC (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or
CC toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism,
CC intoxication by drugs, metal and organic compounds), tumourigenic
CC diseases (e.g. brain tumour), traumatic diseases (e.g. chronic
CC subarachnoidal heamorrhage, and other types of dementia, depression,
CC hyperactive child syndrome (microencephalopathy) and disturbance of
CC consciousness. It is also useful for prevention and treatment of
CC diseases associated with prolactin hypo and hypersecretion
CC respectively, including: hyperprolactinaemia, pituitary adenoma,
CC breast cancer, infertility, impotence and autoimmune disease
CC (hypersecretion disorders), and seminal vesicle hypoplasia,
CC osteoporosis, menopausal syndrome and renal failure (hyposecretion
CC disorders). The 19P2 polypeptide/amide is also useful as a test
CC reagent for study of the prolactin secretory function or as a
CC lactogogue in mammalian farm animals.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 115; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.2e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawyagrgirpvgrf 31

RESULT 15
W95188
ID W95188 standard; peptide; 31 AA.
XX
AC W95188;
XX
DT 10-MAR-1999 (first entry)
XX
DE Bovine pituitary-derived ligand polypeptide fragment.
XX
KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal; bovine.
XX
OS Bos sp.
XX
PN WO9849295-A1.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-JP01923.
XX
PR 28-APR-1997; 97JP-0109974.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fukusumi S, Hinuma S;
XX
DR WPI; 1999-009423/01.
XX
PT New polypeptide ligand for orphan G protein coupled receptors - used
PT for treating disorders of central nervous system, pituitary and
PT pancreas, and for drug screening
XX
PS Example 19; Page 150; 206pp; English.
XX
CC The invention relates to a murine pituitary-derived ligand polypeptide
CC which is a ligand for the G-protein coupled orphan receptor designated
CC GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC the ligand polypeptide encoding DNA are used to produce a recombinant
CC ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC function of the pituitary, central nervous system, pancreas and other
CC tissues and can be used to screen for agents that modulate binding of the
CC polypeptide to the receptor; to quantify the amount of receptor in a
CC sample and to raise antibodies. They may also be used therapeutically,
CC e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC rheumatoid arthritis, epilepsy and many others, also to improve post-
CC operative nutritional status and as vasopressor. Transgenic animals
CC carrying the ligand polypeptide encoding DNA or its mutein are used to
CC study the function of the polypeptide-expressing genes, as models of
CC disease, for drug screening and as source of cell lines. The ligand
CC polypeptide DNA is used as a source of probes and primers; to identify
CC related sequences; in receptor-binding assays; for production of Ab and
CC antisera; in drug development; for gene therapy and to develop transgenic
CC animals. The present sequence represents a bovine genome-derived ligand
CC polypeptide fragment which is similar to the murine ligand-polypeptide.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 115; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.2e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYAGRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawyagrgirpvgrf 31

Search completed: April 17, 2001, 15:38:42

Job time: 349 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:39:47 ; Search time 61.54 Seconds
(without alignments)
6.243 Million cell updates/sec

Title: US-09-446-543A-8
Perfect score: 115
Sequence: 1 TPDINPAWYAGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	115	100.0	20	3	US-09-105-678A-34	Sequence 34, Appl
2	115	100.0	21	3	US-09-105-678A-35	Sequence 35, Appl
3	115	100.0	22	3	US-09-105-678A-36	Sequence 36, Appl
4	115	100.0	31	3	US-09-105-678A-7	Sequence 7, Appli
5	115	100.0	31	3	US-09-105-678A-31	Sequence 31, Appl
6	115	100.0	32	3	US-09-105-678A-32	Sequence 32, Appl
7	115	100.0	33	3	US-09-105-678A-33	Sequence 33, Appl
8	111	96.5	20	3	US-09-105-678A-40	Sequence 40, Appl
9	111	96.5	21	3	US-09-105-678A-41	Sequence 41, Appl
10	111	96.5	22	3	US-09-105-678A-42	Sequence 42, Appl
11	111	96.5	31	3	US-09-105-678A-8	Sequence 8, Appli
12	111	96.5	31	3	US-09-105-678A-37	Sequence 37, Appl
13	111	96.5	31	4	US-09-172-353-4	Sequence 4, Appli
14	111	96.5	32	3	US-09-105-678A-38	Sequence 38, Appl
15	111	96.5	33	3	US-09-105-678A-39	Sequence 39, Appl
16	109	94.8	19	3	US-09-105-678A-30	Sequence 30, Appl
17	109	94.8	20	3	US-09-105-678A-46	Sequence 46, Appl
18	109	94.8	21	3	US-09-105-678A-47	Sequence 47, Appl
19	109	94.8	22	3	US-09-105-678A-48	Sequence 48, Appl
20	109	94.8	31	3	US-09-105-678A-9	Sequence 9, Appli
21	109	94.8	31	3	US-09-105-678A-43	Sequence 43, Appl
22	109	94.8	32	3	US-09-105-678A-44	Sequence 44, Appl
23	109	94.8	33	3	US-09-105-678A-45	Sequence 45, Appl
24	104	90.4	21	3	US-09-105-678A-28	Sequence 28, Appl
25	104	90.4	29	3	US-09-105-678A-29	Sequence 29, Appl
26	50	43.5	349	1	US-08-118-270-71	Sequence 71, Appl
27	50	43.5	349	5	PCT-US93-08528-71	Sequence 71, Appl
28	46	40.0	555	2	US-08-982-232-7	Sequence 7, Appli
29	46	40.0	555	2	US-08-982-232-14	Sequence 14, Appl
30	44	38.3	774	3	US-08-902-632-2	Sequence 2, Appli
31	44	38.3	774	3	US-09-073-354-1	Sequence 1, Appli
32	44	38.3	774	3	US-08-656-005A-1	Sequence 1, Appli
33	44	38.3	774	4	US-09-073-259-1	Sequence 1, Appli
34	44	38.3	774	4	US-09-363-095-1	Sequence 1, Appli
35	41	35.7	70	4	US-08-513-974B-34	Sequence 34, Appl
36	41	35.7	70	4	US-08-513-974B-317	Sequence 317, App
37	41	35.7	141	4	US-08-513-974B-320	Sequence 320, App
38	41	35.7	209	4	US-08-513-974B-321	Sequence 321, App
39	41	35.7	313	2	US-08-446-806-1	Sequence 1, Appli
40	41	35.7	330	2	US-08-815-176-1	Sequence 1, Appli
41	40	34.8	396	1	US-08-229-284A-2	Sequence 2, Appli
42	40	34.8	426	6	5268463-8	Patent No. 5268463
43	40	34.8	428	6	5432081-9	Patent No. 5432081
44	40	34.8	456	6	5432081-7	Patent No. 5432081
45	40	34.8	457	6	5268463-7	Patent No. 5268463

ALIGNMENTS

```
RESULT 1
US-09-105-678A-34
; Sequence 34, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-34


  Query Match           100.0%;  Score 115;  DB 3;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 5.2e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 TPDINPAWYAGRGIRPVGRF 20


RESULT   2
US-09-105-678A-35
; Sequence 35, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-35


  Query Match             100.0%;  Score 115;  DB 3;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 TPDINPAWYAGRGIRPVGRF 20


RESULT   3
US-09-105-678A-36
; Sequence 36, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-36


  Query Match             100.0%;  Score 115;  DB 3;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 TPDINPAWYAGRGIRPVGRF 20


RESULT   4
US-09-105-678A-7
; Sequence 7, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
```

```
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-7


  Query Match           100.0%;  Score 115;  DB 3;  Length 31;
  Best Local Similarity  100.0%;  Pred. No. 8.5e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYAGRGIRPVGRF 31


RESULT   5
US-09-105-678A-31
; Sequence 31, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-31


  Query Match           100.0%;  Score 115;  DB 3;  Length 31;
  Best Local Similarity  100.0%;  Pred. No. 8.5e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYAGRGIRPVGRF 31


RESULT   6
US-09-105-678A-32
; Sequence 32, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-32


  Query Match           100.0%;  Score 115;  DB 3;  Length 32;
  Best Local Similarity  100.0%;  Pred. No. 8.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYAGRGIRPVGRF 31


RESULT   7
US-09-105-678A-33
; Sequence 33, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;       STREET:  130 Water Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/105,678A
;       FILING DATE:  26-JUN-1998
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  JP 172118/1997
;       FILING DATE:  27-JUN-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Conlin, David G.
;       REGISTRATION NUMBER:  27,026
;       REFERENCE/DOCKET NUMBER:  48466-342
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-523-3400
;       TELEFAX:  617-523-6440
;   INFORMATION FOR SEQ ID NO:  33:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  33 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-09-105-678A-33


  Query Match             100.0%;  Score 115;  DB 3;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYAGRGIRPVGRF 31


RESULT   8
US-09-105-678A-40
; Sequence 40, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-40


  Query Match             96.5%;  Score 111;  DB 3;  Length 20;
  Best Local Similarity   95.0%;  Pred. No. 2.2e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          ||||||||| ||||||||||
Db      1 TPDINPAWYTGRGIRPVGRF 20


RESULT   9
US-09-105-678A-41
; Sequence 41, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-41


  Query Match             96.5%;  Score 111;  DB 3;  Length 21;
  Best Local Similarity   95.0%;  Pred. No. 2.4e-11;
```

```
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          |||||||| ||||||||||
Db      1 TPDINPAWYTGRGIRPVGRF 20


RESULT  10
US-09-105-678A-42
; Sequence 42, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-42


  Query Match             96.5%;  Score 111;  DB 3;  Length 22;
  Best Local Similarity   95.0%;  Pred. No. 2.5e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          |||||||| ||||||||||
Db      1 TPDINPAWYTGRGIRPVGRF 20


RESULT  11
US-09-105-678A-8
; Sequence 8, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-8


  Query Match             96.5%;  Score 111;  DB 3;  Length 31;
  Best Local Similarity   95.0%;  Pred. No. 3.6e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          |||||||| ||||||||||
Db     12 TPDINPAWYTGRGIRPVGRF 31


RESULT  12
US-09-105-678A-37
; Sequence 37, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-37


  Query Match             96.5%;  Score 111;  DB 3;  Length 31;
  Best Local Similarity   95.0%;  Pred. No. 3.6e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYAGRGIRPVGRF 20
           |||||||||  ||||||||||
Db      12 TPDINPAWYTGRGIRPVGRF 31


RESULT  13
US-09-172-353-4
; Sequence 4, Application US/09172353
; Patent No. 6197530
; GENERAL INFORMATION:
;  APPLICANT: Stricker-Kongra, Alain
;  APPLICANT: Gu, Wei
;  TITLE OF INVENTION: GPR10 AS A TARGET FOR IDENTIFYING WEIGHT MODULATING COMPOUNDS
;  FILE REFERENCE: 07334/102001
;  CURRENT APPLICATION NUMBER: US/09/172,353
;  CURRENT FILING DATE: 1998-10-14
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 31
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-172-353-4


  Query Match             96.5%;  Score 111;  DB 4;  Length 31;
  Best Local Similarity   95.0%;  Pred. No. 3.6e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYAGRGIRPVGRF 20
           |||||||||  ||||||||||
Db      12 TPDINPAWYTGRGIRPVGRF 31


RESULT  14
US-09-105-678A-38
; Sequence 38, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-38


  Query Match             96.5%;  Score 111;  DB 3;  Length 32;
  Best Local Similarity   95.0%;  Pred. No. 3.7e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYAGRGIRPVGRF 20
           |||||||||  ||||||||||
Db      12 TPDINPAWYTGRGIRPVGRF 31


RESULT  15
US-09-105-678A-39
; Sequence 39, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
```

; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-105-678A-39

Query Match 96.5%; Score 111; DB 3; Length 33;
Best Local Similarity 95.0%; Pred. No. 3.9e-11;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 TPDINPAWYAGRGIRPVGRF 20
           ||||||||| ||||||||||
Db      12 TPDINPAWYTGRGIRPVGRF 31
```

Search completed: April 17, 2001, 15:39:47
Job time: 316 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:48:45 ; Search time 39.1 Seconds
(without alignments)
17.522 Million cell updates/sec

Title: US-09-446-543A-8
Perfect score: 115
Sequence: 1 TPDINPAWYAGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	115	100.0	98	1	PRRP_BOVIN	P81264 bos taurus
2	111	96.5	83	1	PRRP_RAT	P81278 rattus norv
3	109	94.8	87	1	PRRP_HUMAN	P81277 homo sapien
4	46	40.0	424	1	SAMB_SALTY	P23832 salmonella
5	46	40.0	962	1	UVRA_METTH	O26543 methanobact
6	45	39.1	676	1	EXL1_HUMAN	Q92935 homo sapien
7	45	39.1	798	1	UNR_RAT	P18395 rattus norv
8	45	39.1	973	1	UVRA_RHIME	P56899 rhizobium m
9	44	38.3	179	1	RK6_GUITH	O46908 guillardia
10	44	38.3	383	1	CYCR_CHRVI	O82947 chromatium
11	43	37.4	241	1	YG3D_YEAST	P53281 saccharomyc
12	43	37.4	359	1	ALF_HAEIN	P44429 haemophilus
13	43	37.4	476	1	YAAJ_ECOLI	P30143 escherichia
14	43	37.4	511	1	CP12_CANFA	P56592 canis famil
15	43	37.4	960	1	UVRA_TREPA	O83527 treponema p
16	43	37.4	1282	1	TP2M_DICDI	P90520 dictyosteli
17	42	36.5	272	1	TRA2_DROVI	O02008 drosophila
18	42	36.5	332	1	LYTB_MYCLE	Q9x781 mycobacteri
19	42	36.5	374	1	YHHJ_ECOLI	P31993 escherichia
20	42	36.5	453	1	TBB2_GEOCN	P32925 geotrichum
21	42	36.5	719	1	ARP_YEAST	P32770 saccharomyc
22	42	36.5	972	1	UVRA_MYCTU	P94972 mycobacteri
23	41.5	36.1	652	1	TETP_CLOPE	Q46306 clostridium
24	41	35.7	264	1	Y355_BUCAI	P57436 buchnera ap
25	41	35.7	313	1	SPEB_STRCL	P37819 streptomyce
26	41	35.7	322	1	GRP2_MOUSE	O89100 m grb2-rela
27	41	35.7	330	1	GRP2_HUMAN	O75791 h grb2-rela
28	41	35.7	342	1	Y762_METJA	Q58172 methanococc
29	41	35.7	347	1	Y576_METJA	Q57996 methanococc
30	41	35.7	376	1	OPS1_LIMPO	P35360 limulus pol
31	41	35.7	376	1	OPS2_LIMPO	P35361 limulus pol
32	41	35.7	391	1	GAT5_CHICK	P43692 gallus gall
33	41	35.7	423	1	GCRC_MOUSE	P30731 mus musculu
34	41	35.7	546	1	CHOD_STRSQ	P12676 streptomyce
35	41	35.7	620	1	YJB9_YEAST	P47069 saccharomyc
36	41	35.7	652	1	DREB_CHICK	P18302 gallus gall
37	41	35.7	775	1	DPOL_THES9	Q56366 thermococcu
38	41	35.7	943	1	UVRA_HAEIN	P44410 haemophilus
39	41	35.7	986	1	CYGR_ARBPU	P11528 arbacia pun
40	40.5	35.2	265	1	UBIE_RICPR	Q9zcp3 rickettsia
41	40.5	35.2	551	1	ETV4_HUMAN	P43268 homo sapien
42	40.5	35.2	592	1	HIS5_ARATH	Q9sz30 arabidopsis
43	40	34.8	149	1	ENRN_BPT7	P00641 bacteriopha
44	40	34.8	247	1	MER5_ARATH	P24806 arabidopsis
45	40	34.8	424	1	IMPB_SALTY	P18642 salmonella

ALIGNMENTS

RESULT 1
PRRP_BOVIN
ID PRRP_BOVIN STANDARD; PRT; 98 AA.
AC P81264;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE RELEASING PEPTIDE PRRP20].
GN PRH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.
RC TISSUE=Brain;
RX MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RL Nature 393:272-276(1998).
CC -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AB015417; BAA29025.1; -.
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 22
FT PEPTIDE 23 53 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT PEPTIDE 33 53 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 53 53 AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ SEQUENCE 98 AA; 10544 MW; 08AC35A13B0FA908 CRC64;

Query Match 100.0%; Score 115; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 7.2e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TPDINPAWYAGRGIRPVGRF 20
           ||||||||||||||||||||
Db      34 TPDINPAWYAGRGIRPVGRF 53
```

```
RESULT   2
PRRP_RAT
ID   PRRP_RAT       STANDARD;      PRT;    83 AA.
AC   P81278;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015418; BAA29026.1; -.
KW   Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     21       BY SIMILARITY.
FT   PEPTIDE      22     52       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      33     52       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      52     52       AMIDATION (G-53 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   83 AA;  9215 MW;  D0C75A264EEE4F29 CRC64;


  Query Match             96.5%;  Score 111;  DB 1;  Length 83;
  Best Local Similarity   95.0%;  Pred. No. 2.5e-10;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          |||||||||| ||||||||||
Db     33 TPDINPAWYTGRGIRPVGRF 52


RESULT   3
PRRP_HUMAN
ID   PRRP_HUMAN     STANDARD;      PRT;    87 AA.
AC   P81277;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015419; BAA29027.1; -.
DR   MIM; 602663; -.
KW   Hormone; Amidation; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   PEPTIDE      23     53       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      34     53       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      53     53       AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   87 AA;  9639 MW;  229A2F3F50CF981B CRC64;


  Query Match             94.8%;  Score 109;  DB 1;  Length 87;
  Best Local Similarity   95.0%;  Pred. No. 5.3e-10;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          |||||||||| ||||||||||
Db     34 TPDINPAWYASRGIRPVGRF 53


RESULT   4
SAMB_SALTY
ID   SAMB_SALTY     STANDARD;      PRT;   424 AA.
AC   P23832;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   SAMB PROTEIN.
GN   SAMB.
OS   Salmonella typhimurium.
OG   Plasmid 60-MDa cryptic.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=91123176; PubMed=1991707;
RA   Nohmi T., Hakura A., Nakai Y., Watanabe M., Murayama S.Y.,
RA   Sofuni T.;
RT   "Salmonella typhimurium has two homologous but different umuDC
RT   operons: cloning of a new umuDC-like operon (samAB) present in a
RT   60-megadalton cryptic plasmid of S. typhimurium.";
RL   J. Bacteriol. 173:1051-1063(1991).
CC   -!- FUNCTION: INVOLVED IN UV PROTECTION AND MUTATION.
CC   -!- SIMILARITY: BELONGS TO THE IMPB/MUCB/SAMB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; D90202; BAA14226.1; -.
DR   PIR; B38176; B38176.
DR   InterPro; IPR001126; -.
DR   Pfam; PF00817; IMS; 1.
KW   Plasmid; SOS mutagenesis; DNA repair.
SQ   SEQUENCE   424 AA;  47727 MW;  FF8C47476CC58A2B CRC64;


  Query Match             40.0%;  Score 46;  DB 1;  Length 424;
  Best Local Similarity   53.3%;  Pred. No. 9.4;
  Matches    8;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYAGRGIRP 16
           |     |:||||| |
Db     384 PGKGKIWFAGRGIAP 398


RESULT   5
UVRA_METTH
ID   UVRA_METTH     STANDARD;      PRT;   962 AA.
AC   O26543;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR MTH443.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000828; AAB84949.1; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      38     45       ATP (POTENTIAL).
FT   NP_BIND     649    656       ATP (POTENTIAL).
FT   ZN_FING     748    774       C4-TYPE.
SQ   SEQUENCE   962 AA;  108395 MW;  2C0EF7FC41CCD060 CRC64;


  Query Match             40.0%;  Score 46;  DB 1;  Length 962;
  Best Local Similarity   37.8%;  Pred. No. 21;
  Matches   14;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy       1 TPDINPAWYAG----------------RGIRPVGRF 20
           ||  ||| | |                  || || |||
Db     704 TPRSNPATYTGVFTHIRELFAQTPEARKRGYRP-GRF 739


RESULT   6
EXL1_HUMAN
ID   EXL1_HUMAN     STANDARD;      PRT;   676 AA.
AC   Q92935;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   EXOSTOSIN-LIKE 1 (EXOSTOSIN-L) (MULTIPLE EXOSTOSIS-LIKE PROTEIN).
GN   EXTL1 OR EXTL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97189339; PubMed=9037597;
RA   Wise C.A., Clines G.A., Massa H., Trask B.J., Lovett M.;
RT   "Identification and localization of the gene for EXTL, a third member
RT   of the multiple exostoses gene family.";
RL   Genome Res. 7:10-16(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Xu L., Deng H.X., Xia J.H., Pan Q., Liu C.Y.;
RT   "Mutations of the EXT genes in hereditary multiple exostoses in
RT   Chinese.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Wuyts W., Spieker N., Van Roy N., De Paepe A., De Boulle K.,
RA   Willems P.J., Van Hul W., Versteeg R., Speleman F.;
RT   "Refined physical mapping and genomic structure of the EXTL1 gene.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: APPEARS TO BE A TUMOR SUPPRESSOR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC
CC       RETICULUM (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U67191; AAC51141.1; -.
DR   EMBL; AF083633; AAD02840.1; -.
DR   EMBL; AF083623; AAD02840.1; JOINED.
DR   EMBL; AF083624; AAD02840.1; JOINED.
DR   EMBL; AF083625; AAD02840.1; JOINED.
DR   EMBL; AF083626; AAD02840.1; JOINED.
DR   EMBL; AF083627; AAD02840.1; JOINED.
DR   EMBL; AF083628; AAD02840.1; JOINED.
DR   EMBL; AF083629; AAD02840.1; JOINED.
DR   EMBL; AF083630; AAD02840.1; JOINED.
DR   EMBL; AF083631; AAD02840.1; JOINED.
DR   EMBL; AF083632; AAD02840.1; JOINED.
DR   EMBL; AF153980; AAF73172.1; -.
DR   EMBL; AF151391; AAF73172.1; JOINED.
DR   MIM; 601738; -.
KW   Anti-oncogene; Multigene family; Transmembrane; Signal-anchor.
FT   TRANSMEM     10     30       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
```

```
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   676 AA;  74673 MW;  B5E006A8762E5633 CRC64;


  Query Match             39.1%;  Score 45;  DB 1;  Length 676;
  Best Local Similarity   45.0%;  Pred. No. 21;
  Matches    9;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPVGRF 20
          :|   | :|  :| || |||
Db    400 SPQDFPPFYYLQQGSRPEGRF 419


RESULT   7
UNR_RAT
ID   UNR_RAT        STANDARD;      PRT;   798 AA.
AC   P18395;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UNR PROTEIN.
GN   UNR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=90370473; PubMed=2204029;
RA   Jeffers M., Paciucci R., Pellicer A.;
RT   "Characterization of unr; a gene closely linked to N-ras.";
RL   Nucleic Acids Res. 18:4891-4899(1990).
CC   -!- FUNCTION: RNA-BINDING PROTEIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52311; CAA36549.1; -.
DR   PIR; S11210; S11210.
DR   HSSP; P15277; 1MJC.
DR   InterPro; IPR002059; -.
DR   Pfam; PF00313; CSD; 8.
DR   PROSITE; PS00352; COLD_SHOCK; 4.
KW   RNA-binding; Repeat.
FT   DOMAIN       26     87       CSD 1.
FT   DOMAIN      136    179       CSD 2 (INCOMPLETE).
FT   DOMAIN      186    245       CSD 3.
FT   DOMAIN      297    337       CSD 4 (INCOMPLETE).
FT   DOMAIN      349    410       CSD 5.
FT   DOMAIN      447    507       CSD 6.
FT   DOMAIN      519    579       CSD 7.
FT   DOMAIN      610    670       CSD 8.
FT   DOMAIN      674    735       CSD 9.
SQ   SEQUENCE   798 AA;  88894 MW;  F484B3FA8B0995A4 CRC64;


  Query Match             39.1%;  Score 45;  DB 1;  Length 798;
  Best Local Similarity   47.1%;  Pred. No. 24;
  Matches    8;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRPV 17
          | : ||  |:|: |||:
Db    589 TEEANPTIYSGKVIRPL 605
```

```
RESULT   8
UVRA_RHIME
ID   UVRA_RHIME     STANDARD;      PRT;   973 AA.
AC   P56899;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RA   Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
RA   Boistard P., Gouzy J., Kahn D., Thebault P., Goffeau A.,
RA   Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
RA   Vandenbol M., Puehler A., Becker A., Weidner S.;
RL   Submitted (MAY-2000) to the SWISS-PROT data bank.
RN   [2]
RP   SEQUENCE OF 1-140 FROM N.A.
RC   STRAIN=2021;
RX   MEDLINE=99430868; PubMed=10503543;
RA   Tapias A., Barbe J.;
RT   "Regulation of divergent transcription from the uvrA-ssb promoters in
RT   Sinorhizobium meliloti.";
RL   Mol. Gen. Genet. 262:121-130(1999).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF125162; AAF03210.1; -.
DR   InterPro; IPR001617; -.
DR   PROSITE; PS00211; ABC_TRANSPORTER; PARTIAL.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      34     41       ATP (POTENTIAL).
FT   NP_BIND     662    669       ATP (POTENTIAL).
FT   ZN_FING     761    787       C4-TYPE.
FT   CONFLICT     19     19       G -> A (IN REF. 2).
FT   CONFLICT     67     67       F -> S (IN REF. 2).
SQ   SEQUENCE   973 AA;  107191 MW;  3E1A8B14527A47FE CRC64;


  Query Match             39.1%;  Score 45;  DB 1;  Length 973;
  Best Local Similarity   37.8%;  Pred. No. 29;
  Matches   14;  Conservative    2;  Mismatches    3;  Indels   18;  Gaps    3;

Qy      1 TPDINPA----------WYAG------RGIRPVGRF 20
          ||  |||          |:||      || :| |||
Db    717 TPRSNPATYTGAFTPIRDWFAGLPEAKARGYQP-GRF 752


RESULT   9
RK6_GUITH
ID   RK6_GUITH      STANDARD;      PRT;   179 AA.
AC   O46908;
```

```
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CHLOROPLAST 50S RIBOSOMAL PROTEIN L6.
GN   RPL6.
OS   Guillardia theta (Cryptomonas phi).
OG   Chloroplast.
OC   Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX   NCBI_TaxID=55529;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97283757; PubMed=9137835;
RA   Wang S.L., Liu X.-Q., Douglas S.E.;
RT   "The large ribosomal protein gene cluster of a cryptomonad plastid:
RT   gene organization, sequence and evolutionary implications.";
RL   Biochem. Mol. Biol. Int. 41:1035-1044(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99128221; PubMed=9929392;
RA   Douglas S.E., Penny S.L.;
RT   "The plastid genome of the cryptophyte alga, Guillardia theta:
RT   complete sequence and conserved synteny groups confirm its common
RT   ancestry with red algae.";
RL   J. Mol. Evol. 48:236-244(1999).
CC   -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
CC       LOCATED AT THE AMINOACYL-TRNA BINDING SITE OF THE
CC       PEPTIDYLTRANSFERASE CENTER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L6P FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF041468; AAC35717.1; -.
DR   InterPro; IPR000702; -.
DR   InterPro; IPR002358; -.
DR   Pfam; PF00347; Ribosomal_L6; 1.
DR   PRINTS; PR00059; RIBOSOMALL6.
DR   PROSITE; PS00525; RIBOSOMAL_L6_1; 1.
KW   Ribosomal protein; Chloroplast.
SQ   SEQUENCE   179 AA;  19527 MW;  8B4C0BDB0152AD24 CRC64;
```

```
  Query Match             38.3%;  Score 44;  DB 1;  Length 179;
  Best Local Similarity   60.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       6 PAWYAGRGIRPVGRF 20
           |  | |:||| || |
Db     154 PEPYKGKGIRYVGEF 168
```

```
RESULT  10
CYCR_CHRVI
ID   CYCR_CHRVI     STANDARD;      PRT;   383 AA.
AC   O82947;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PHOTOSYNTHETIC REACTION CENTER CYTOCHROME C SUBUNIT PRECURSOR.
GN   PUFC.
OS   Chromatium vinosum.
OC   Bacteria; Proteobacteria; gamma subdivision; Chromatiaceae;
OC   Allochromatium.
OX   NCBI_TaxID=1049;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D;
RA   Corson G.E., Nagashima K.V., Matsuura K., Sakuragi Y., Ruwanthi W.,
RA   Qin H., Allen R., Knaff D.B.;
RT   "Primary structure of genes encoding light-harvesting and reaction
RT   center proteins from Chromatium vinosum.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE REACTION CENTER OF PURPLE BACTERIA CONTAIN A
CC       TIGHTLY BOUND CYTOCHROME MOLECULE WHICH RE-REDUCES THE PHOTO
CC       OXIDIZED PRIMARY ELECTRON DONOR.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC       ANCHOR (BY SIMILARITY).
CC   -!- PTM: BINDS FOUR HEME GROUPS PER MOLECULE.
CC   -!- SIMILARITY: HIGH, WITH OTHER PHOTOSYNTHETIC REACTION CENTER
CC       CYTOCHROME C SUBUNITS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB011811; BAA32742.1; -.
DR   InterPro; IPR000345; -.
DR   PROSITE; PS00190; CYTOCHROME_C; 4.
KW   Electron transport; Photosynthesis; Reaction center; Heme;
KW   Membrane; Lipoprotein; Duplication; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    383       PHOTOSYNTHETIC REACTION CENTER CYTOCHROME
FT                                 C SUBUNIT.
FT   LIPID         23     23       N-ACYL DIGLYCERIDE (BY SIMILARITY).
FT   BINDING      107    107       HEME 1 (COVALENT) (BY SIMILARITY).
FT   BINDING      110    110       HEME 1 (COVALENT) (BY SIMILARITY).
FT   METAL        111    111       IRON 1 (HEME AXIAL LIGAND)
FT                                 (BY SIMILARITY).
FT   BINDING      152    152       HEME 2 (COVALENT) (BY SIMILARITY).
FT   BINDING      155    155       HEME 2 (COVALENT) (BY SIMILARITY).
FT   METAL        156    156       IRON 2 (HEME AXIAL LIGAND)
FT                                 (BY SIMILARITY).
FT   BINDING      247    247       HEME 3 (COVALENT) (BY SIMILARITY).
FT   BINDING      250    250       HEME 3 (COVALENT) (BY SIMILARITY).
FT   METAL        251    251       IRON 3 (HEME AXIAL LIGAND)
FT                                 (BY SIMILARITY).
FT   BINDING      307    307       HEME 4 (COVALENT) (BY SIMILARITY).
FT   BINDING      310    310       HEME 4 (COVALENT) (BY SIMILARITY).
FT   METAL        311    311       IRON 4 (HEME AXIAL LIGAND)
FT                                 (BY SIMILARITY).
SQ   SEQUENCE   383 AA;  41522 MW;  96BCD91FF1B9AE7E CRC64;
```

```
  Query Match             38.3%;  Score 44;  DB 1;  Length 383;
  Best Local Similarity   42.1%;  Pred. No. 17;
  Matches    8;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYAGRGIRPVGR 19
           ||    |||| | :| : :
Db     263 TPQRTTAWYAIRHVRDINQ 281
```

```
RESULT  11
YG3D_YEAST
ID   YG3D_YEAST     STANDARD;      PRT;   241 AA.
AC   P53281;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 26.1 KDA PROTEIN IN PAS5-CBF2 INTERGENIC REGION.
GN   YGR136W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RA   van Dyck L., Skala J., de Wergifosse P., Purnelle B., Talla E.,
RA   Nawrocki A., del Bino S., Goffeau A.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z72921; CAA97149.1; -.
DR   HSSP; P29355; 1SEM.
DR   SGD; S0003368; YGR136W.
DR   InterPro; IPR001452; -.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PROSITE; PS50002; SH3; 1.
KW   Hypothetical protein; SH3 domain.
FT   DOMAIN       53    112       SH3.
SQ   SEQUENCE   241 AA;  26139 MW;  5F0B1361AF84AA79 CRC64;


  Query Match             37.4%;  Score 43;  DB 1;  Length 241;
  Best Local Similarity   41.2%;  Pred. No. 16;
  Matches    7;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 INPAWYAGRGIRPVGRF 20
           |:| || |:    :| |
Db      86 ISPDWYRGKSNNKIGIF 102


RESULT  12
ALF_HAEIN
ID   ALF_HAEIN      STANDARD;      PRT;   359 AA.
AC   P44429;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   FRUCTOSE-BISPHOSPHATE ALDOLASE (EC 4.1.2.13).
GN   FBA OR HI0524.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: D-FRUCTOSE 1,6-BISPHOSPHATE = GLYCERONE-
CC       PHOSPHATE + D-GLYCERALDEHYDE 3-PHOSPHATE.
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
CC   -!- PATHWAY: SIXTH STEP IN GLYCOLYSIS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO CLASS II FRUCTOSE-BISPHOSPHATE ALDOLASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32734; AAC22182.1; -.
DR   HSSP; P11604; 1DOS.
DR   TIGR; HI0524; -.
DR   InterPro; IPR000771; -.
DR   Pfam; PF01116; F_bP_aldolase; 1.
DR   PROSITE; PS00602; ALDOLASE_CLASS_II_1; 1.
DR   PROSITE; PS00806; ALDOLASE_CLASS_II_2; 1.
KW   Lyase; Glycolysis; Zinc.
FT   METAL        108    108       ZINC (BY SIMILARITY).
FT   METAL        111    111       ZINC (BY SIMILARITY).
SQ   SEQUENCE   359 AA;  39339 MW;  1EDDFCDD0B69E32C CRC64;


  Query Match             37.4%;  Score 43;  DB 1;  Length 359;
  Best Local Similarity   70.0%;  Pred. No. 23;
  Matches    7;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 AWYAGRGIRP 16
           |:|||:||:|
Db      67 AFYAGKGIKP 76


RESULT  13
YAAJ_ECOLI
ID   YAAJ_ECOLI     STANDARD;      PRT;   476 AA.
AC   P30143;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 51.7 KDA PROTEIN IN THRC-TALB INTERGENIC REGION (ORF8).
GN   YAAJ.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92334977; PubMed=1630901;
RA   Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA   Isono K., Mizobuchi K., Nakata A.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of
RT   the 0-2.4 min region.";
RL   Nucleic Acids Res. 20:3305-3308(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:ALANINE SYMPORTER FAMILY
CC       (SAF). STRONG, TO H.INFLUENZAE HI0183.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; D10483; -; NOT_ANNOTATED_CDS.
DR   EMBL; AE000111; AAC73118.1; -.
DR   EcoGene; EG11555; yaaJ.
DR   InterPro; IPR001463; -.
DR   Pfam; PF01235; Na_Ala_symp; 1.
DR   PRINTS; PR00175; NAALASMPORT.
DR   PROSITE; PS00873; NA_ALANINE_SYMP; 1.
KW   Hypothetical protein; Transmembrane; Inner membrane; Transport;
KW   Symport.
FT   TRANSMEM      4     24       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    141    161       POTENTIAL.
FT   TRANSMEM    174    194       POTENTIAL.
FT   TRANSMEM    207    227       POTENTIAL.
FT   TRANSMEM    233    253       POTENTIAL.
FT   TRANSMEM    300    320       POTENTIAL.
FT   TRANSMEM    351    371       POTENTIAL.
FT   TRANSMEM    391    411       POTENTIAL.
FT   TRANSMEM    414    434       POTENTIAL.
SQ   SEQUENCE   476 AA;  51662 MW;  2F6EB2E12E126E63 CRC64;


  Query Match             37.4%;  Score 43;  DB 1;  Length 476;
  Best Local Similarity   44.4%;  Pred. No. 30;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    6;  Gaps    1;

Qy       3 DIN------PAWYAGRGI 14
             |:|      ||||  ||:
Db     120 DVNGQFRGGPAWYMARGL 137


RESULT  14
CP12_CANFA
ID   CP12_CANFA     STANDARD;      PRT;   511 AA.
AC   P56592;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CYTOCHROME P450 1A2 (EC 1.14.14.1) (CYPIA2) (DAH2) (CYTOCHROME P450-
DE   D2).
GN   CYP1A2.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE OF 9-511 FROM N.A.
RC   STRAIN=BEAGLE; TISSUE=Liver;
RX   MEDLINE=91042464; PubMed=2122230;
RA   Uchida T., Komori M., Kitada M., Kamataki T.;
RT   "Isolation of cDNAs coding for three different forms of liver
RT   microsomal cytochrome P-450 from polychlorinated biphenyl-treated
RT   beagle dogs.";
RL   Mol. Pharmacol. 38:644-651(1990).
RN   [2]
RP   SEQUENCE OF 1-16.
RC   STRAIN=BEAGLE; TISSUE=Liver;
RX   MEDLINE=89087526; PubMed=2910310;
RA   Ohta K., Motoya M., Komori M., Miura T., Kitada M., Kamataki T.;
RT   "A novel form of cytochrome P-450 in beagle dogs. P-450-D3 is a low
RT   spin form of cytochrome P-450 but with catalytic and structural
RT   properties similar to P-450d.";
RL   Biochem. Pharmacol. 38:91-96(1989).
CC   -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC       MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
CC       NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC       OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
CC       ACIDS, AND XENOBIOTICS.
CC   -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC       OXIDIZED FLAVOPROTEIN + H(2)O.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC   -!- TISSUE SPECIFICITY: CONSTITUTIVELY EXPRESSED IN LIVER.
CC   -!- INDUCTION: BY POLYCHLORINATED BIPHENYL (PCB) IN LIVER AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   InterPro; IPR001128; -.
DR   InterPro; IPR002401; -.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PRINTS; PR00463; EP450I.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   INIT_MET      0      0
FT   BINDING     453    453       HEME (BY SIMILARITY).
SQ   SEQUENCE   511 AA;  57505 MW;  200904C54F4B3CE7 CRC64;


  Query Match             37.4%;  Score 43;  DB 1;  Length 511;
  Best Local Similarity   58.3%;  Pred. No. 32;
  Matches    7;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYAGR 12
           :||  | | |||
Db     121 SPDSGPVWAAGR 132


RESULT  15
UVRA_TREPA
ID   UVRA_TREPA     STANDARD;      PRT;   960 AA.
AC   O83527;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR TP0514.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NICHOLS;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001227; AAC65502.1; -.
DR   TIGR; TP0514; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
```

```
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      35     42       ATP (POTENTIAL).
FT   NP_BIND     657    664       ATP (POTENTIAL).
FT   ZN_FING     270    297       C4-TYPE.
FT   ZN_FING     756    782       C4-TYPE.
SQ   SEQUENCE   960 AA;  106010 MW;  32F78624B19F7ABF CRC64;


  Query Match             37.4%;  Score 43;  DB 1;  Length 960;
  Best Local Similarity   35.1%;  Pred. No. 59;
  Matches   13;  Conservative    1;  Mismatches    5;  Indels   18;  Gaps    2;

Qy       1 TPDINPAWYAG----------------RGIRPVGRF 20
           ||  ||| | |                ||  :| |||
Db     712 TPRSNPATYVGVFTDIRMLFSQVPEAKMRGYKP-GRF 747


Search completed: April 17, 2001, 15:48:47
Job time: 535 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:45:51 ; Search time 70.08 Seconds
(without alignments)
19.613 Million cell updates/sec

Title: US-09-446-543A-8
Perfect score: 115
Sequence: 1 TPDINPAWYAGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	51	44.3	430	1	B69009	conserved hypothet
2	50	43.5	527	2	T33175	hypothetical prote
3	50	43.5	790	2	T47959	hypothetical prote
4	48.5	42.2	664	2	F83376	conserved hypothet
5	46	40.0	333	2	H82852	hydroxybenzoate oc
6	46	40.0	424	2	B38176	samB protein - Sal
7	46	40.0	962	2	H69157	excinuclease ABC c
8	45.5	39.6	779	2	T49717	related to BCS1 pr
9	45	39.1	284	2	F71015	hypothetical prote
10	45	39.1	767	2	T21969	hypothetical prote
11	45	39.1	798	2	S11210	probable unr prote
12	44	38.3	390	2	G82844	cysteine synthase
13	44	38.3	1670	2	S71551	DNA-directed DNA p
14	43.5	37.8	506	2	H83396	probable aldehyde
15	43.5	37.8	506	2	F83142	probable aldehyde
16	43.5	37.8	1501	2	T45623	hypothetical prote
17	43	37.4	220	2	C83292	probable glutathio
18	43	37.4	241	1	S64445	hypothetical prote
19	43	37.4	276	2	D70817	hypothetical prote
20	43	37.4	309	2	T32376	hypothetical prote
21	43	37.4	359	2	C64074	fructose-bisphosph
22	43	37.4	399	2	T30222	sensory protein ki
23	43	37.4	476	2	G64720	probable amino aci
24	43	37.4	503	2	A82193	Sun/nucleolar prot
25	43	37.4	511	2	B37222	cytochrome P450 1A
26	43	37.4	548	2	T47548	hypothetical prote
27	43	37.4	960	2	A71315	excinuclease ABC c
28	43	37.4	1282	2	T30577	DNA topoisomerase
29	43	37.4	1292	2	T31462	probable magnesium
30	42	36.5	143	2	T36978	probable transposa
31	42	36.5	284	2	A75117	hypothetical prote
32	42	36.5	311	2	S66600	cytochrome-c oxida
33	42	36.5	375	2	S47704	hypothetical 41.1K
34	42	36.5	428	2	F81694	pyruvate dehydroge
35	42	36.5	453	2	S18597	tubulin beta chain
36	42	36.5	468	2	C83160	nitrite extrusion
37	42	36.5	719	2	S61046	ARP1 protein - yea
38	42	36.5	940	2	A82329	excinuclease ABC,
39	42	36.5	972	2	A70619	excinuclease ABC c
40	42	36.5	1296	2	T16859	hypothetical prote
41	41.5	36.1	165	2	S72776	B1496_F1_41 protei
42	41.5	36.1	345	2	D84012	N-acetylglutamate
43	41.5	36.1	443	2	T21499	hypothetical prote
44	41.5	36.1	652	1	S41522	tetracycline resis
45	41.5	36.1	7962	2	I38346	elastic titin - hu

ALIGNMENTS

RESULT 1
B69009
conserved hypothetical protein MTH1070 - Methanobacterium thermoautotrophicum (strain
C;Species: Methanobacterium thermoautotrophicum
C;Date: 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change 23-Jul-1999
C;Accession: B69009
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: B69009
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-430 <MTH>
A;Cross-references: GB:AE000877; GB:AE000666; NID:g2622157; PIDN:AAB85559.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1070
C;Superfamily: conserved hypothetical protein MTH1070

Query Match 44.3%; Score 51; DB 1; Length 430;
Best Local Similarity 69.2%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       3 DINPAWYAGRGIR 15
           |||| | |||  |
Db     191 DINPEWVAGRACR 203
```

RESULT 2
T33175
hypothetical protein C24G6.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33175
R;Greco, T.; Bradshaw, H.; Keppler, D.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid C24G6.
A;Reference number: Z21298
A;Accession: T33175
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-527 <GRE>
A;Cross-references: EMBL:AF067936; PIDN:AAC19213.1; GSPDB:GN00023; CESP:C24G6.6
A;Experimental source: strain Bristol N2; clone C24G6
C;Genetics:
A;Gene: CESP:C24G6.6
A;Map position: 5
A;Introns: 20/3; 77/1; 129/2; 208/3; 470/2

```
  Query Match             43.5%;  Score 50;  DB 2;  Length 527;
  Best Local Similarity   66.7%;  Pred. No. 6.1;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYAGRG 13
           |::  |||||||
Db     370 PNVLSAWYAGRG 381


RESULT   3
T47959
hypothetical protein F15G16.60 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47959
R;De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quet
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z24480
A;Accession: T47959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-790 <DEH>
A;Cross-references: EMBL:AL132959
A;Experimental source: cultivar Columbia; BAC clone F15G16
C;Genetics:
A;Map position: 3
A;Introns: 39/1; 678/2; 698/3; 773/2
A;Note: F15G16.60


  Query Match             43.5%;  Score 50;  DB 2;  Length 790;
  Best Local Similarity   47.4%;  Pred. No. 9.2;
  Matches    9;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYAGRGIRPVGRF 20
           |  ||  |  ||::| ||:
Db     366 PPHNPRTYGSRGLQPHGRW 384


RESULT   4
F83376
conserved hypothetical protein PA2151 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83376
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337
A;Accession: F83376
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-664 <STO>
A;Cross-references: GB:AE004642; GB:AE004091; NID:g9948163; PIDN:AAG05539.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2151


  Query Match             42.2%;  Score 48.5;  DB 2;  Length 664;
  Best Local Similarity   62.5%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       1 TPDINPAWYAGRGIRP 16
           |||||| |:  |  ||
Db     478 TPDINP-WFLQRSGRP 492


RESULT   5
H82852
hydroxybenzoate octaprenyltransferase XF0068 [imported] - Xylella fastidiosa (strain
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
C;Accession: H82852
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H82852
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-333 <SIM>
A;Cross-references: GB:AE003860; GB:AE003849; NID:g9104830; PIDN:AAF82881.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0068
C;Superfamily: 4-hydroxybenzoate octaprenyltransferase


  Query Match             40.0%;  Score 46;  DB 2;  Length 333;
  Best Local Similarity   53.3%;  Pred. No. 16;
  Matches    8;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 INPAWYAGRGIRPVG 18
           ::| |   || ||||
Db      54 LDPYWKLARGDRPVG 68


RESULT   6
B38176
samB protein - Salmonella typhimurium
C;Species: Salmonella typhimurium
C;Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 20-Jun-2000
C;Accession: B38176
R;Nohmi, T.; Hakura, A.; Nakai, Y.; Watanabe, M.; Murayama, S.Y.; Sofuni, T.
J. Bacteriol. 173, 1051-1063, 1991
A;Title: Salmonella typhimurium has two homologous but different umuDC operons: cloni
A;Reference number: A38176; MUID:91123176
A;Accession: B38176
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-424 <NOH>
A;Cross-references: GB:D90202; NID:g217087; PIDN:BAA14226.1; PID:g217089
A;Experimental source: strain LT2
C;Genetics:
A;Gene: samB
C;Function:
A;Description: restores UV mutability; involved in mutagenesis
C;Superfamily: umuC protein
C;Keywords: DNA repair; induced mutagenesis; SOS mutagenesis


  Query Match             40.0%;  Score 46;  DB 2;  Length 424;
  Best Local Similarity   53.3%;  Pred. No. 20;
  Matches    8;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYAGRGIRP 16
```

```
          |     |:|||||  |
Db    384 PGKGKIWFAGRGIAP 398


RESULT   7
H69157
excinuclease ABC chain A - Methanobacterium thermoautotrophicum (strain Delta H)
N;Alternate names: uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 12-Nov-1999
C;Accession: H69157
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514
A;Accession: H69157
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-962 <MTH>
A;Cross-references: GB:AE000828; GB:AE000666; NID:g2621504; PIDN:AAB84949.1; PID:g262150
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH443
A;Start codon: TTG
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; P-loop
F;38-45/Region: nucleotide-binding motif A (P-loop)
F;632-915/Domain: ATP-binding cassette homology <ABCE>
F;649-656/Region: nucleotide-binding motif A (P-loop)

  Query Match             40.0%;  Score 46;  DB 2;  Length 962;
  Best Local Similarity   37.8%;  Pred. No. 46;
  Matches   14;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      1 TPDINPAWYAG----------------RGIRPVGRF 20
          ||  ||| | |                 || || |||
Db    704 TPRSNPATYTGVFTHIRELFAQTPEARKRGYRP-GRF 739


RESULT   8
T49717
related to BCS1 protein precursor [imported] - Neurospora crassa
N;Alternate names: protein B23L21.300
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49717
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49717
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-779 <SCH>
A;Cross-references: EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.300
A;Experimental source: BAC clone B23L21; strain OR74A
C;Genetics:
A;Gene: NCSP:B23L21.300
A;Map position: 6

  Query Match             39.6%;  Score 45.5;  DB 2;  Length 779;
  Best Local Similarity   58.8%;  Pred. No. 44;
  Matches   10;  Conservative    1;  Mismatches    3;  Indels    3;  Gaps    1;

Qy      1 TPDINPA---WYAGRGI 14
          |  :|||   ||| |||
Db    286 TDYLNPATRRWYANRGI 302


RESULT   9
F71015
hypothetical protein PH1420 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: F71015
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A;Reference number: A71000; MUID:98344137
A;Accession: F71015
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-284 <KAW>
A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30526.1; PID:g3257843
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa
C;Genetics:
A;Gene: PH1420
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1420

  Query Match             39.1%;  Score 45;  DB 2;  Length 284;
  Best Local Similarity   50.0%;  Pred. No. 19;
  Matches    9;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      2 PDINPAWYAGRGIRPVGR 19
          | | | :|| ||:  :||
Db    217 PYIEPTFYALRGLELLGR 234


RESULT  10
T21969
hypothetical protein F38E11.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21969
R;Matthews, P.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19495
A;Accession: T21969
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-767 <WIL>
A;Cross-references: EMBL:Z68342; PIDN:CAA92775.1; GSPDB:GN00022; CESP:F38E11.7
A;Experimental source: clone F38E11
C;Genetics:
A;Gene: CESP:F38E11.7
A;Map position: 4
A;Introns: 50/2; 118/1; 139/2; 189/3; 226/1; 248/1; 287/2; 375/2; 432/3; 465/3; 548/1

  Query Match             39.1%;  Score 45;  DB 2;  Length 767;
  Best Local Similarity   75.0%;  Pred. No. 52;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 PDINPAWY 9
          ||: ||||
Db    747 PDVKPAWY 754


RESULT  11
S11210
probable unr protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C;Accession: S11210
R;Jeffers, M.; Paciucci, R.; Pellicer, A.
Nucleic Acids Res. 18, 4891-4899, 1990
A;Title: Characterization of unr; a gene closely linked to N-ras.
A;Reference number: S11210; MUID:90370473
A;Accession: S11210
```

```
A;Molecule type: mRNA
A;Residues: 1-798 <JEF>
A;Cross-references: EMBL:X52311; NID:g57454; PIDN:CAA36549.1; PID:g57455
C;Keywords: DNA binding


  Query Match             39.1%;  Score 45;  DB 2;  Length 798;
  Best Local Similarity   47.1%;  Pred. No. 54;
  Matches    8;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYAGRGIRPV 17
           | : ||  |:|: |||:
Db     589 TEEANPTIYSGKVIRPL 605


RESULT  12
G82844
cysteine synthase XF0128 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: G82844
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: G82844
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-390 <SIM>
A;Cross-references: GB:AE003866; GB:AE003849; NID:g9104906; PIDN:AAF82941.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0128


  Query Match             38.3%;  Score 44;  DB 2;  Length 390;
  Best Local Similarity   70.0%;  Pred. No. 37;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 NPAWYAGRGI 14
           :|||||  ||
Db     354 SPAWYAAHGI 363


RESULT  13
S71551
DNA-directed DNA polymerase (EC 2.7.7.7) KOD, intein containing precursor - Pyrococcus s
N;Contains: DNA endonuclease (EC 3.1.-.-) PI-I; DNA endonuclease (EC 3.1.-.-) PI-II; DNA
C;Species: Pyrococcus sp.
A;Variety: strain KOD1
C;Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 20-Apr-2000
C;Accession: S71551
R;Kakihara, H.; Takagi, M.; Imanaka, T.
submitted to the EMBL Data Library, March 1994
A;Description: Cloning and analysis of the DNA polymerase gene from a new hyperthermophi
A;Reference number: S71551
A;Accession: S71551
A;Molecule type: DNA
A;Residues: 1-1670 <KAK>
A;Cross-references: EMBL:D29671
C;Function: <NUCL>
A;Description: as DNA-directed DNA polymerase, catalyzes the polymerization of DNA at
A;Note: DNA-directed DNA polymerase KOD
C;Function: <EN1>
A;Description: as DNA endonuclease PI-I, catalyzes the hydrolysis of internal phospho
A;Note: DNA endonuclease PI-PspI
C;Function: <EN2>
A;Description: as DNA endonuclease PI-II, catalyzes the hydrolysis of internal phosph
A;Note: DNA endonuclease PI-PspII
C;Superfamily: DNA-directed DNA polymerase KOD
C;Keywords: endonuclease; hydrolase; nucleotidyltransferase; protein splicing
F;1-406,767-851,1388-1670/Product: DNA-directed DNA polymerase KOD #status predicted
F;1-406/Domain: DNA-directed DNA polymerase KOD extein 1 #status predicted <XT1>
F;407-766/Product: DNA endonuclease PI-I (pol KOD intein 1) #status predicted <MAT2>
F;767-851/Domain: DNA-directed DNA polymerase KOD extein 2 #status predicted <XT2>
F;852-1387/Product: DNA endonuclease PI-II (pol KOD intein 2) #status predicted <MAT3
F;1388-1670/Domain: DNA-directed DNA polymerase KOD extein 3 #status predicted <XT3>
F;406-767/Cross-link: peptide (Arg-Ser) #status predicted
F;851-1388/Cross-link: peptide (Asn-Ser) #status predicted


  Query Match             38.3%;  Score 44;  DB 2;  Length 1670;
  Best Local Similarity   38.9%;  Pred. No. 1.6e+02;
  Matches    7;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 DINPAWYAGRGIRPVGRF 20
           :| |||  |  :  | |:
Db     294 EITPAWETGENLERVARY 311


RESULT  14
H83396
probable aldehyde dehydrogenase PA1984 [imported] - Pseudomonas aeruginosa (strain PA
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83396
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: H83396
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-506 <STO>
A;Cross-references: GB:AE004625; GB:AE004091; NID:g9947983; PIDN:AAG05372.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1984
C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology


  Query Match             37.8%;  Score 43.5;  DB 2;  Length 506;
  Best Local Similarity   57.9%;  Pred. No. 58;
  Matches   11;  Conservative    2;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1 TPDINPAWYAGRGIRPVGR 19
           | ||| |:  ||||:  ||
Db     436 TRDINRAYRMGRGIK-AGR 453


RESULT  15
F83142
probable aldehyde dehydrogenase PA4022 [imported] - Pseudomonas aeruginosa (strain PA
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83142
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
```

Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337
A;Accession: F83142
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-506 <STO>
A;Cross-references: GB:AE004819; GB:AE004091; NID:g9950214; PIDN:AAG07409.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4022
C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 37.8%; Score 43.5; DB 2; Length 506;
Best Local Similarity 57.9%; Pred. No. 58;
Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

```
Qy      1 TPDINPAWYAGRGIRPVGR 19
          | ||| |:  ||||:  ||
Db    436 TRDINRAYRMGRGIK-AGR 453
```

Search completed: April 17, 2001, 15:45:53
Job time: 601 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:48:03 ; Search time 115.07 Seconds
(without alignments)
20.372 Million cell updates/sec

Title: US-09-446-543A-8
Perfect score: 115
Sequence: 1 TPDINPAWYAGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	90	78.3	117	13	Q9W624	Q9w624 carassius a
2	57	49.6	54	4	Q9UJF9	Q9ujf9 homo sapien
3	57	49.6	465	4	O60687	O60687 homo sapien
4	51	44.3	430	1	O27142	O27142 methanobact
5	50	43.5	527	5	O76383	O76383 caenorhabdi
6	50	43.5	790	10	Q9M371	Q9m371 arabidopsis
7	48	41.7	343	3	O74569	O74569 coprinus ci
8	46.5	40.4	333	2	Q9RJI0	Q9rji0 streptomyce
9	46	40.0	333	2	Q9PH76	Q9ph76 xylella fas
10	46	40.0	555	3	O00050	O00050 aspergillus
11	46	40.0	820	5	Q9NE93	Q9ne93 leishmania
12	45.5	39.6	779	3	Q9P5J9	Q9p5j9 neurospora
13	45	39.1	284	1	O50128	O50128 pyrococcus
14	45	39.1	767	5	Q20170	Q20170 caenorhabdi
15	44	38.3	153	9	Q9T133	Q9t133 bacteriopha
16	44	38.3	390	2	Q9PH18	Q9ph18 xylella fas
17	44	38.3	398	10	Q9SJR5	Q9sjr5 arabidopsis
18	44	38.3	419	4	Q9Y276	Q9y276 homo sapien
19	44	38.3	540	10	Q9LGZ0	Q9lgz0 oryza sativ
20	43.5	37.8	506	2	Q9ZAA1	Q9zaa1 pseudomonas
21	43.5	37.8	1501	10	Q9SD86	Q9sd86 arabidopsis
22	43	37.4	66	5	O44613	O44613 caenorhabdi
23	43	37.4	223	2	O85605	O85605 prevotella
24	43	37.4	267	5	Q9V3E5	Q9v3e5 drosophila
25	43	37.4	276	2	O53929	O53929 mycobacteri
26	43	37.4	306	2	Q9L4D4	Q914d4 xanthomonas
27	43	37.4	309	5	O17234	O17234 caenorhabdi
28	43	37.4	351	2	O07306	O07306 prochloroco
29	43	37.4	399	2	Q54293	Q54293 streptomyce
30	43	37.4	429	2	Q9L0X9	Q910x9 streptomyce
31	43	37.4	503	2	Q9KRY1	Q9kry1 vibrio chol
32	43	37.4	548	10	Q9LFA0	Q91fa0 arabidopsis
33	43	37.4	627	2	Q9L8G6	Q918g6 clostridium
34	43	37.4	630	4	Q9NXS8	Q9nxs8 homo sapien
35	43	37.4	938	10	Q9ZVE3	Q9zve3 arabidopsis
36	43	37.4	1292	2	Q9ZGE5	Q9zge5 heliobacill
37	42.5	37.0	243	2	Q9R6H5	Q9r6h5 agrobacteri
38	42.5	37.0	506	2	Q9KH02	Q9kh02 azotobacter
39	42	36.5	72	8	Q9TD69	Q9td69 mytilus cal
40	42	36.5	143	2	Q9RI97	Q9ri97 streptomyce
41	42	36.5	207	5	O76756	O76756 apis mellif
42	42	36.5	284	1	Q9V0Q3	Q9v0q3 pyrococcus
43	42	36.5	360	3	Q9P8B3	Q9p8b3 agaricus bi
44	42	36.5	428	2	Q9RDL4	Q9rdl4 streptomyce
45	42	36.5	428	2	Q9PKE7	Q9pke7 chlamydia m

ALIGNMENTS

```
RESULT   1
Q9W624
ID   Q9W624      PRELIMINARY;      PRT;   117 AA.
AC   Q9W624;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   C-RF AMIDE PRECURSOR.
OS   Carassius auratus (Goldfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Cyprininae; Carassius.
OX   NCBI_TaxID=7957;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Satake H., Minakata H., Fujimoto M.;
RT   "Carassius RFamide (C-RF amide).";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB020024; BAA76662.1; -.
SQ   SEQUENCE   117 AA;  12879 MW;  D5DC4CB22038C2B0 CRC64;

  Query Match             78.3%;  Score 90;  DB 13;  Length 117;
  Best Local Similarity   65.0%;  Pred. No. 2.5e-06;
  Matches   13;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYAGRGIRPVGRF 20
           :|:|:| || |||:||:|||
Db      56 SPEIDPFWYVGRGVRPIGRF 75


RESULT   2
Q9UJF9
ID   Q9UJF9      PRELIMINARY;      PRT;    54 AA.
AC   Q9UJF9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DJ479J7.3 (SUSHI-REPEAT PROTEIN (SRPUL)) (FRAGMENT).
GN   DJ479J7.3.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lawlor S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035608; CAB55682.1; -.
FT   NON_TER      54     54
SQ   SEQUENCE   54 AA;  6110 MW;  E2F3C39F7B961A9F CRC64;


  Query Match             49.6%;  Score 57;  DB 4;  Length 54;
  Best Local Similarity   56.2%;  Pred. No. 0.11;
  Matches    9;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRP 16
          || : | |||| |  |
Db     18 TPAVTPTWYAGSGYYP 33


RESULT   3
O60687
ID   O60687      PRELIMINARY;      PRT;   465 AA.
AC   O60687;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   SUSHI-REPEAT PROTEIN.
GN   SRPUL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kurosawa H., Inukai T., Inaba T., Goi K., Chang K.-S., Sinjyo T.,
RA   Rakestraw K.M., Naeve C.W., Look T.A.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF060567; AAC15765.1; -.
DR   INTERPRO; IPR000436; -.
DR   INTERPRO; IPR001128; -.
DR   PFAM; PF00084; sushi; 3.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
SQ   SEQUENCE   465 AA;  52971 MW;  4D752B187FF3EFB8 CRC64;


  Query Match             49.6%;  Score 57;  DB 4;  Length 465;
  Best Local Similarity   56.2%;  Pred. No. 1.1;
  Matches    9;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYAGRGIRP 16
          || : | |||| |  |
Db     18 TPAVTPTWYAGSGYYP 33


RESULT   4
O27142
ID   O27142      PRELIMINARY;      PRT;   430 AA.
AC   O27142;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CONSERVED PROTEIN.
GN   MTH1070.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanobacterium.
OX   NCBI_TaxID=2166;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000877; AAB85559.1; -.
DR   INTERPRO; IPR002510; -.
DR   PFAM; PF01523; PmbA_TldD; 1.
SQ   SEQUENCE   430 AA;  46062 MW;  C4FAE47C111749E6 CRC64;


  Query Match             44.3%;  Score 51;  DB 1;  Length 430;
  Best Local Similarity   69.2%;  Pred. No. 8;
  Matches    9;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 DINPAWYAGRGIR 15
          |||| | |||  |
Db    191 DINPEWVAGRACR 203


RESULT   5
O76383
ID   O76383      PRELIMINARY;      PRT;   527 AA.
AC   O76383;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   C24G6.6 PROTEIN.
GN   C24G6.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Greco T., Bradshaw H., Keppler D.;
RT   "The sequence of C. elegans cosmid C24G6.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF067936; AAC19213.1; -.
DR   INTERPRO; IPR000205; -.
DR   INTERPRO; IPR002937; -.
DR   PFAM; PF01593; Amino_oxidase; 1.
SQ   SEQUENCE   527 AA;  59805 MW;  9FBB1FB84437C5CB CRC64;
```

```
  Query Match             43.5%;  Score 50;  DB 5;  Length 527;
  Best Local Similarity   66.7%;  Pred. No. 14;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYAGRG 13
           |::  ||||||
Db     370 PNVLSAWYAGRG 381


RESULT   6
Q9M371
ID   Q9M371      PRELIMINARY;      PRT;   790 AA.
AC   Q9M371;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 87.4 KDA PROTEIN.
GN   F15G16.60.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   De Haan M., Maarse A.C., Grivell L.A., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL132959; CAB71097.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   790 AA;  87375 MW;  B222724B75690F30 CRC64;


  Query Match             43.5%;  Score 50;  DB 10;  Length 790;
  Best Local Similarity   47.4%;  Pred. No. 22;
  Matches    9;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYAGRGIRPVGRF 20
           |  ||  |  ||::| ||:
Db     366 PPHNPRTYGSRGLQPHGRW 384


RESULT   7
O74569
ID   O74569      PRELIMINARY;      PRT;   343 AA.
AC   O74569;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   RAH1.
GN   RAH1.
OS   Coprinus cinereus (Inky cap fungus).
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Agaricales;
OC   Coprinaceae; Coprinus.
OX   NCBI_TaxID=5346;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OKAYAMA-7;
RX   MEDLINE=97174112; PubMed=9021132;
RA   Yeager Stassen N., Logsdon J.M. Jr., Vora G.J., Offenberg H.H.,
RA   Palmer J.D., Zolan M.E.;
RT   "Isolation and characterization of rad51 orthologs from Coprinus
RT   cinereus and Lycopersicon esculentum, and phylogenetic analysis of
RT   eukaryotic recA homologs.";
RL   Curr. Genet. 31:144-157(1997).
DR   EMBL; U21905; AAC23703.1; -.
DR   INTERPRO; IPR000445; -.
DR   INTERPRO; IPR001553; -.
DR   PFAM; PF00633; HHH; 1.
SQ   SEQUENCE   343 AA;  36789 MW;  EA8AA6BB5B83D226 CRC64;


  Query Match             41.7%;  Score 48;  DB 3;  Length 343;
  Best Local Similarity   50.0%;  Pred. No. 18;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYAGRGIRPVG 18
           |||  |  |||   :|:|
Db     273 TPDAAPGPYAGNEKKPIG 290


RESULT   8
Q9RJI0
ID   Q9RJI0      PRELIMINARY;      PRT;   333 AA.
AC   Q9RJI0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 36.3 KDA PROTEIN.
GN   SCF73.06C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL121746; CAB57411.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   333 AA;  36312 MW;  884FF2861837FD76 CRC64;


  Query Match             40.4%;  Score 46.5;  DB 2;  Length 333;
  Best Local Similarity   35.5%;  Pred. No. 29;
  Matches   11;  Conservative    2;  Mismatches    7;  Indels   11;  Gaps    1;

Qy       1 TPDINPAW----------YAGRGIRPVGRF 20
           ||   |||           |  |:|| ||:
Db      19 TPWEEPAWRTEAIGWIEARLAAHGLRPTGRW 49


RESULT   9
Q9PH76
ID   Q9PH76      PRELIMINARY;      PRT;   333 AA.
AC   Q9PH76;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYDROXYBENZOATE OCTAPRENYLTRANSFERASE.
GN   XF0068.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
```

```
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-157(2000).
DR   EMBL; AE003860; AAF82881.1; -.
DR   INTERPRO; IPR000537; -.
DR   PFAM; PF01040; UbiA; 1.
DR   PROSITE; PS00943; UBIA; UNKNOWN_1.
SQ   SEQUENCE   333 AA;  37931 MW;  ECF3F4716C962B95 CRC64;


  Query Match             40.0%;  Score 46;  DB 2;  Length 333;
  Best Local Similarity   53.3%;  Pred. No. 35;
  Matches    8;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 INPAWYAGRGIRPVG 18
          ::| |   || ||||
Db     54 LDPYWKLARGDRPVG 68


RESULT  10
O00050
ID   O00050      PRELIMINARY;      PRT;   555 AA.
AC   O00050;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   TRANSPOSASE.
OS   Aspergillus awamori.
OC   Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;
OC   anamorphic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=105351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UVK143F DERIVED FROM NRRL3112; TRANSPOSON=VADER;
RX   MEDLINE=97156905; PubMed=9003286;
RA   Nyyssonen E., Amutan M., Enfield L., Stubbs J., Dunn-Coleman N.S.;
RT   "The transposable element Tanl of Aspergillus niger var. awamori, a
RT   new member of the Fotl family.";
RL   Mol. Gen. Genet. 253:50-56(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UVK143F DERIVED FROM NRRL3112; TRANSPOSON=VADER;
RX   MEDLINE=96207472; PubMed=8625427;
RA   Amutan M., Nyyssonen E., Stubbs J., Diaz-Torres M.R., Dunn-Coleman N.;
RT   "Identification and cloning of a mobile transposon from Aspergillus
RT   niger var. awamori.";
RL   Curr. Genet. 29:468-473(1996).
DR   EMBL; U58946; AAC49623.1; -.
SQ   SEQUENCE   555 AA;  62978 MW;  70904D2EE009EA33 CRC64;


  Query Match             40.0%;  Score 46;  DB 3;  Length 555;
  Best Local Similarity   57.1%;  Pred. No. 60;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      7 AWYAGRGIRPVGRF 20
          ||| |: | |  ||
Db    237 AWYEGQSIPPTWRF 250


RESULT  11
Q9NE93
ID   Q9NE93      PRELIMINARY;      PRT;   820 AA.
AC   Q9NE93;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 86.4 KDA PROTEIN.
GN   L787.05.
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRIEDLIN;
RA   Robben J., Aert R., Volckaert G., Ivens A.C., Quail M.,
RA   Rajandream M.A., Barrell B.G.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRIEDLIN;
RA   Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA   Smith D.F.;
RT   "A physical map of the Leishmania major Friedlin genome.";
RL   Genome Res. 8:135-145(1998).
DR   EMBL; AL163492; CAB86689.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   820 AA;  86388 MW;  F4CB0D9FDE666817 CRC64;


  Query Match             40.0%;  Score 46;  DB 5;  Length 820;
  Best Local Similarity   47.1%;  Pred. No. 91;
  Matches    8;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 PDINPAWYAGRGIRPVG 18
          |  :|| |:| | :|:|
Db    406 PFTSPALYSGNGSQPLG 422


RESULT  12
Q9P5J9
ID   Q9P5J9      PRELIMINARY;      PRT;   779 AA.
AC   Q9P5J9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   RELATED TO BCS1 PROTEIN PRECURSOR.
GN   B23L21.300.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; AL356172; CAB91698.1; -.
SQ   SEQUENCE   779 AA;  85796 MW;  CA7891402DFFBE30 CRC64;


  Query Match             39.6%;  Score 45.5;  DB 3;  Length 779;
  Best Local Similarity   58.8%;  Pred. No. 1e+02;
  Matches   10;  Conservative    1;  Mismatches    3;  Indels    3;  Gaps    1;

Qy       1 TPDINPA---WYAGRGI 14
           |  :|||   ||| |||
Db     286 TDYLNPATRRWYANRGI 302


RESULT  13
O50128
ID   O50128      PRELIMINARY;      PRT;   284 AA.
AC   O50128;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 32.3 KDA PROTEIN PH1420.
GN   PH1420.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000006; BAA30526.1; -.
DR   INTERPRO; IPR001066; -.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   284 AA;  32319 MW;  8E0E7BC3711D3815 CRC64;


  Query Match             39.1%;  Score 45;  DB 1;  Length 284;
  Best Local Similarity   50.0%;  Pred. No. 42;
  Matches    9;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYAGRGIRPVGR 19
           | | | :|| ||:  :||
Db     217 PYIEPTFYALRGLELLGR 234


RESULT  14
Q20170
ID   Q20170      PRELIMINARY;      PRT;   767 AA.
AC   Q20170;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   F38E11.7 PROTEIN.
GN   F38E11.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matthews P.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z68342; CAA92775.1; -.
DR   INTERPRO; IPR000595; -.
DR   INTERPRO; IPR001622; -.
DR   INTERPRO; IPR002025; -.
DR   PFAM; PF00027; cNMP_binding; 1.
DR   PFAM; PF00914; CNG_membrane; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; UNKNOWN_1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
SQ   SEQUENCE   767 AA;  89988 MW;  F7ECF69DBBEAACF3 CRC64;


  Query Match             39.1%;  Score 45;  DB 5;  Length 767;
  Best Local Similarity   75.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 PDINPAWY 9
           ||: ||||
Db     747 PDVKPAWY 754


RESULT  15
Q9T133
ID   Q9T133      PRELIMINARY;      PRT;   153 AA.
AC   Q9T133;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENDONUCLEASE.
GN   3.
OS   Bacteriophage phi-YeO3-12.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae;
OC   T7-like phages.
OX   NCBI_TaxID=110457;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pajunen M.I., Kiljunen S.J., Skurnik M.;
RT   "Complete genomic sequence of the lytic bacteriophage phi-YeO3-12 of
RT   Yersinia enterocolitica serotype O:3.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ251805; CAB63604.1; -.
KW   Endonuclease.
SQ   SEQUENCE   153 AA;  17640 MW;  211571BBDE6C641D CRC64;


  Query Match             38.3%;  Score 44;  DB 9;  Length 153;
  Best Local Similarity   66.7%;  Pred. No. 30;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       9 YAGRGIRPVGRF 20
           || ||:| || |
Db       5 YAARGVRKVGAF 16


Search completed: April 17, 2001, 15:48:06
Job time: 559 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:45:53 ; Search time 70.08 Seconds
(without alignments)
30.400 Million cell updates/sec

Title: US-09-446-543A-47
Perfect score: 174
Sequence: 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	35.1	1236	2	T50904	Mg protoporphyrin
2	58	33.3	1292	2	T31462	probable magnesium
3	56	32.2	962	2	H69157	excinuclease ABC c
4	53	30.5	798	2	S11210	probable unr prote
5	53	30.5	1415	2	C83070	conserved hypothet
6	52	29.9	940	2	A82329	excinuclease ABC,
7	52	29.9	972	2	A70619	excinuclease ABC c
8	51.5	29.6	503	2	A82193	Sun/nucleolar prot
9	51	29.3	294	2	T21075	hypothetical prote
10	51	29.3	943	2	D64057	excinuclease ABC c
11	50.5	29.0	176	2	S67150	hypothetical prote
12	50.5	29.0	548	2	T47548	hypothetical prote
13	50	28.7	482	1	S40887	RVS167 protein - y
14	50	28.7	940	1	BVECUA	excinuclease ABC c
15	50	28.7	953	2	D71645	excinuclease ABC c
16	50	28.7	965	2	C82560	excinuclease ABC s
17	49.5	28.4	375	2	S47704	hypothetical 41.1K
18	49	28.2	128	2	S76955	hypothetical prote
19	49	28.2	220	2	C83292	probable glutathio
20	49	28.2	772	2	T07958	protoporphyrin IX
21	49	28.2	945	2	E83117	excinuclease ABC s
22	49	28.2	952	1	T46550	excinuclease ABC c
23	49	28.2	970	2	S77349	excinuclease ABC c
24	49	28.2	1014	2	T36031	excinuclease ABC c
25	49	28.2	1193	2	T50729	magnesium-protopor
26	49	28.2	1331	2	S75000	protoporphyrin IX
27	49	28.2	1379	2	S37310	protoporphyrin IX
28	49	28.2	1380	2	S64721	protoporphyrin IX
29	49	28.2	1381	2	S71288	protoporphyrin IX
30	49	28.2	1382	2	T01789	protoporphyrin IX
31	49	28.2	1383	2	T07126	magnesium chelatas
32	49	28.2	1670	2	S71551	DNA-directed DNA p
33	48.5	27.9	664	2	F83376	conserved hypothet
34	48	27.6	157	2	A81811	hypothetical prote
35	48	27.6	455	2	D70885	probable aldC prot
36	48	27.6	498	2	T09021	beta-glucosidase h
37	48	27.6	517	2	T09022	beta-glucosidase h
38	48	27.6	719	2	S61046	ARP1 protein - yea
39	48	27.6	790	2	T47959	hypothetical prote
40	48	27.6	798	2	S29815	N-ras upstream pro
41	48	27.6	948	2	B81883	excinuclease ABC c
42	48	27.6	949	2	A81138	excinuclease ABC c
43	48	27.6	960	2	A71315	excinuclease ABC c
44	48	27.6	1194	2	D49851	magnesium-protopor
45	47.5	27.3	501	2	I61512	TNF receptor assoc

ALIGNMENTS

RESULT 1
T50904
Mg protoporphyrin methyl transferase [imported] - Rubrivivax gelatinosus
C;Species: Rubrivivax gelatinosus
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 02-Sep-2000
C;Accession: T50904
R;Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.
submitted to the EMBL Data Library, November 1999
A;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photos
A;Reference number: Z25270
A;Accession: T50904
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1236 <NAG>
A;Cross-references: EMBL:AB034704; PIDN:BAA94057.1
A;Experimental source: strain IL144
C;Genetics:
A;Gene: bchH
C;Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase

Query Match 35.1%; Score 61; DB 2; Length 1236;
Best Local Similarity 37.5%; Pred. No. 1.3;
Matches 12; Conservative 6; Mismatches 8; Indels 6; Gaps 1;

```
Qy       3 AHQHSMETRTPDINPAWYTG------RGIRPV 28
           : |  ::||||  :|| || |       |:| :
Db    1112 SEQVALETRTRMLNPKWYEGMLEHGYEGVRQI 1143
```

RESULT 2
T31462
probable magnesium chelatase (EC 4.99.1.-) chain H BchH - Heliobacillus mobilis
C;Species: Heliobacillus mobilis
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 15-Sep-2000
C;Accession: T31462
R;Xiong, J.; Inoue, K.; Bauer, C.E.
Proc. Natl. Acad. Sci. U.S.A. 95, 14851-14856, 1998
A;Title: Tracking molecular evolution of photosynthesis by characterization of a majo
A;Reference number: Z21036; MUID:99061957
A;Accession: T31462
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1292 <XIO>
A;Cross-references: EMBL:AF080002; NID:g3820536; PID:g3820560; PIDN:AAC84033.1
C;Genetics:
A;Gene: bchH
C;Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase
C;Keywords: lyase

Query Match 33.3%; Score 58; DB 2; Length 1292;

```
  Best Local Similarity   42.3%;  Pred. No. 3.8;
  Matches   11;  Conservative    3;  Mismatches    6;  Indels    6;  Gaps    1;

Qy       9 ETRTPDINPAWYTG------RGIRPV 28
           ||||  :|| || |       |:| :
Db    1176 ETRTKTLNPKWYEGMLKHGYEGVREI 1201


RESULT   3
H69157
excinuclease ABC chain A - Methanobacterium thermoautotrophicum (strain Delta H)
N;Alternate names: uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 12-Nov-1999
C;Accession: H69157
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514
A;Accession: H69157
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-962 <MTH>
A;Cross-references: GB:AE000828; GB:AE000666; NID:g2621504; PIDN:AAB84949.1; PID:g262150
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH443
A;Start codon: TTG
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; P-loop
F;38-45/Region: nucleotide-binding motif A (P-loop)
F;632-915/Domain: ATP-binding cassette homology <ABCE>
F;649-656/Region: nucleotide-binding motif A (P-loop)


  Query Match             32.2%;  Score 56;  DB 2;  Length 962;
  Best Local Similarity   42.1%;  Pred. No. 5.4;
  Matches   16;  Conservative    0;  Mismatches    4;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                 || || |||
Db     703 RTPRSNPATYTGVFTHIRELFAQTPEARKRGYRP-GRF 739


RESULT   4
S11210
probable unr protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C;Accession: S11210
R;Jeffers, M.; Paciucci, R.; Pellicer, A.
Nucleic Acids Res. 18, 4891-4899, 1990
A;Title: Characterization of unr; a gene closely linked to N-ras.
A;Reference number: S11210; MUID:90370473
A;Accession: S11210
A;Molecule type: mRNA
A;Residues: 1-798 <JEF>
A;Cross-references: EMBL:X52311; NID:g57454; PIDN:CAA36549.1; PID:g57455
C;Keywords: DNA binding


  Query Match             30.5%;  Score 53;  DB 2;  Length 798;
  Best Local Similarity   43.5%;  Pred. No. 12;
  Matches   10;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       6 HSMETRTPDINPAWYTGRGIRPV 28
           ||:   | : ||  |:|: |||:
Db     583 HSVNGITEEANPTIYSGKVIRPL 605


RESULT   5
C83070
conserved hypothetical protein PA4601 [imported] - Pseudomonas aeruginosa (strain PAO
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: C83070
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: C83070
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1415 <STO>
A;Cross-references: GB:AE004874; GB:AE004091; NID:g9950849; PIDN:AAG07989.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4601


  Query Match             30.5%;  Score 53;  DB 2;  Length 1415;
  Best Local Similarity   37.5%;  Pred. No. 22;
  Matches    9;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMETRTPDINPAWYTGRG 24
           ::| || :    ||: | | |  |
Db     330 AQARQHHLLGLLPDLEPGWLTDAG 353


RESULT   6
A82329
excinuclease ABC, chain A VC0394 [imported] - Vibrio cholerae (group O1 strain N16961
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C;Accession: A82329
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: A82329
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-940 <HEI>
A;Cross-references: GB:AE004127; GB:AE003852; NID:g9654808; PIDN:AAF93567.1; GSPDB:GN
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0394
A;Map position: 1
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology


  Query Match             29.9%;  Score 52;  DB 2;  Length 940;
  Best Local Similarity   39.5%;  Pred. No. 20;
  Matches   15;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                 || :| |||
Db     694 RTPRSNPATYTGIFTPIRELFAGTQESRSRGYQP-GRF 730


RESULT   7
A70619
excinuclease ABC chain A - Mycobacterium tuberculosis (strain H37RV)
N;Alternate names: uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70619
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
```

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: A70619
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-972 <COL>
A;Cross-references: GB:Z85982; GB:AL123456; NID:g3261718; PIDN:CAB06633.1; PID:g1838989
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: uvrA
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; hydrolase
F;32-39/Region: nucleotide-binding motif A (P-loop)
F;637-920/Domain: ATP-binding cassette homology <ABCE>
F;654-661/Region: nucleotide-binding motif A (P-loop)

Query Match 29.9%; Score 52; DB 2; Length 972;
Best Local Similarity 39.5%; Pred. No. 21;
Matches 15; Conservative 1; Mismatches 4; Indels 18; Gaps 2;

```
Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                || :| |||
Db     708 RTPRSNPATYTGVFDKIRTLFAATTEAKVRGYQP-GRF 744
```

RESULT 8
A82193
Sun/nucleolar protein family protein VC1502 [imported] - Vibrio cholerae (group O1 strai
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C;Accession: A82193
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: A82193
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-503 <HEI>
A;Cross-references: GB:AE004228; GB:AE003852; NID:g9655997; PIDN:AAF94657.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC1502
A;Map position: 1

Query Match 29.6%; Score 51.5; DB 2; Length 503;
Best Local Similarity 36.7%; Pred. No. 12;
Matches 11; Conservative 5; Mismatches 13; Indels 1; Gaps 1;

```
Qy       1 SRAHQHSMETRTPDINPAWYTGRGIRPVGR 30
           | :  ||:|  |      |: || :|| |:
Db     435 SSSASHSVELDTTQAR-EWFMGRDVRPEGQ 463
```

RESULT 9
T21075
hypothetical protein F19H6.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T21075; T21124
R;McMurray, A.
submitted to the EMBL Data Library, August 1995
A;Reference number: Z19368
A;Accession: T21075
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-294 <WIL>
A;Cross-references: EMBL:Z50873; PIDN:CAA90762.1; GSPDB:GN00028; CESP:F19H6.1
A;Experimental source: clone F17E5
R;McMurray, A.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19378
A;Accession: T21124
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-294 <WI2>
A;Cross-references: EMBL:Z68115; PIDN:CAA92169.1; GSPDB:GN00028; CESP:F19H6.1
A;Experimental source: clone F19H6
C;Genetics:
A;Gene: CESP:F19H6.1
A;Map position: X
A;Introns: 55/3; 85/2; 113/3; 150/1; 192/3; 266/2
C;Superfamily: kinase-related transforming protein; protein kinase homology

Query Match 29.3%; Score 51; DB 2; Length 294;
Best Local Similarity 44.8%; Pred. No. 7.9;
Matches 13; Conservative 2; Mismatches 12; Indels 2; Gaps 1;

```
Qy       3 AHQHSMETRTPDINPA--WYTGRGIRPVG 29
           || ||      || ||  : || ||  :|
Db     139 AHMHSKRIMHRDIKPANVFITGNGIVKLG 167
```

RESULT 10
D64057
excinuclease ABC chain A - Haemophilus influenzae
N;Alternate names: uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 12-Nov-1999
C;Accession: D64057; JC5157
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: D64057
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-943 <TIGR>
A;Cross-references: GB:U32711; GB:L42023; NID:g1573214; PIDN:AAC21915.1; PID:g1573215
A;Experimental source: strain Rd KW20
R;de la Morena, M.L.; Hendrixson, D.R.; St Geme III, J.W.
Gene 177, 23-28, 1996
A;Title: Isolation and characterization of the Haemophilus influenzae uvrA gene.
A;Reference number: JC5157; MUID:97080495
A;Accession: JC5157
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-162,'L',164-235,'D',237-424,'K',426-462,'M',464-513,'Q',515-660,'T',662
A;Cross-references: GB:U33877; NID:g1144488; PIDN:AAC44592.1; PID:g1144489
A;Experimental source: strain N187
C;Genetics:
A;Gene: uvrA
C;Function:
A;Description: has ATPase and DNA binding activity; involved in DNA repair
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; P-loop
F;31-38/Region: nucleotide-binding motif A (P-loop)
F;623-907/Domain: ATP-binding cassette homology <ABCE>
F;640-647/Region: nucleotide-binding motif A (P-loop)

Query Match 29.3%; Score 51; DB 2; Length 943;
Best Local Similarity 39.5%; Pred. No. 28;

```
  Matches   15;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                 ||  | |||
Db     695 RTPRSNPATYTGLFTPIRELFAGVPEARARGYNP-GRF 731


RESULT  11
S67150
hypothetical protein YOR253w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein O5315
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 05-Nov-1999
C;Accession: S67150
R;Jauniaux, J.C.; Poirey, R.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67143
A;Accession: S67150
A;Molecule type: DNA
A;Residues: 1-176 <JAU>
A;Cross-references: EMBL:Z75161; NID:g1420572; PID:e252411; PID:g1420573; GSPDB:GN00015;
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YOR253w
A;Map position: 15R


  Query Match             29.0%;  Score 50.5;  DB 2;  Length 176;
  Best Local Similarity   31.2%;  Pred. No. 5.4;
  Matches   10;  Conservative    4;  Mismatches   15;  Indels    3;  Gaps    1;

Qy       1 SRAHQHSMETRTP---DINPAWYTGRGIRPVG 29
           |  |||::   |   |:   |:   |   ||
Db     120 SECHQHNVFVYLPAVDDLTKQWFIAHGFEQVG 151


RESULT  12
T47548
hypothetical protein F8J2.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47548
R;Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;
Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24458
A;Accession: T47548
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-548 <NYA>
A;Cross-references: EMBL:AL132969
A;Experimental source: cultivar Columbia; BAC clone F8J2
C;Genetics:
A;Map position: 3
A;Introns: 78/1; 143/1; 242/2; 377/3; 428/2; 447/3; 470/3; 502/3; 532/3
A;Note: F8J2.80


  Query Match             29.0%;  Score 50.5;  DB 2;  Length 548;
  Best Local Similarity   35.5%;  Pred. No. 18;
  Matches   11;  Conservative    3;  Mismatches   10;  Indels    7;  Gaps    1;

Qy       1 SRAHQ-------HSMETRTPDINPAWYTGRG 24
           |  ||       | ::   |   |:|| |||
Db     118 SLLHQSPMHFLHHPLQHSFPHHQPSWYWGRG 148


RESULT  13
S40887
RVS167 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YDR388w
C;Species: Saccharomyces cerevisiae
C;Date: 31-Mar-1992 #sequence_revision 06-Feb-1995 #text_change 21-Jul-2000
C;Accession: S40887; S69672
R;Bauer, F.; Urdaci, M.; Aigle, M.; Crouzet, M.
Mol. Cell. Biol. 13, 5070-5084, 1993
A;Title: Alteration of a yeast SH3 protein leads to conditional viability with defect
A;Reference number: S40887; MUID:93330299
A;Accession: S40887
A;Molecule type: DNA
A;Residues: 1-482 <BAU>
A;Cross-references: EMBL:M92092; NID:g172615; PIDN:AAA35051.1; PID:g172616
R;Dietrich, F.S.
submitted to the EMBL Data Library, July 1995
A;Description: The sequence of S. cerevisiae cosmids 9481, 9509, 9926, 9461, and lamb
A;Reference number: S69665
A;Accession: S69672
A;Molecule type: DNA
A;Residues: 1-482 <DIE>
A;Cross-references: EMBL:U32274; NID:g927313; PIDN:AAB64830.1; PID:g927321; GSPDB:GN0
C;Genetics:
A;Gene: SGD:RVS167; MIPS:YDR388w
A;Cross-references: SGD:S0002796; MIPS:YDR388w
A;Map position: 4R
C;Superfamily: RVS167 protein; RVS161 protein homology; SH3 homology
C;Keywords: transmembrane protein
F;4-270/Domain: RVS161 protein homology <RVS>
F;292-422/Region: alanine/glycine/proline-rich
F;428-477/Domain: SH3 homology <SH3>


  Query Match             28.7%;  Score 50;  DB 1;  Length 482;
  Best Local Similarity   50.0%;  Pred. No. 19;
  Matches   11;  Conservative    3;  Mismatches    2;  Indels    6;  Gaps    2;

Qy      11 RTPDINPAWYTGR-----GIRP 27
           ||||:|  |:|||     .|: |
Db     453 RTPDVN-EWWTGRYNGQQGVFP 473


RESULT  14
BVECUA
excinuclease ABC chain A - Escherichia coli
N;Alternate names: uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Escherichia coli
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 12-Nov-1999
C;Accession: A23869; I78011; A65214
R;Husain, I.; Van Houten, B.; Thomas, D.C.; Sancar, A.
J. Biol. Chem. 261, 4895-4901, 1986
A;Title: Sequences of Escherichia coli uvrA gene and protein reveal two potential ATP
A;Reference number: A23869; MUID:86168204
A;Accession: A23869
A;Molecule type: DNA
A;Residues: 1-940 <HUS>
A;Cross-references: GB:M13495; NID:g148164; PIDN:AAA24754.1; PID:g148165
R;Sancar, A.; Sancar, G.B.; Rupp, W.D.; Little, J.W.; Mount, D.W.
Nature 298, 96-98, 1982
A;Title: LexA protein inhibits transcription of the E. coli uvrA gene in vitro.
A;Reference number: I58044; MUID:82220077
A;Accession: I78011
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-14 <RES>
A;Cross-references: GB:J01721; NID:g148161; PIDN:AAA24753.1; PID:g551847
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: A65214
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-940 <BLAT>
A;Cross-references: GB:AE000479; GB:U00096; NID:g2367340; PIDN:AAC77028.1; PID:g23673
A;Experimental source: strain K-12, substrain MG1655
```

C;Comment: This protein is an ATPase and a DNA-binding protein that preferentially binds
e, an ATP-dependent DNA repair enzyme that catalyzes the excision reaction of UV-damaged
C;Genetics:
A;Gene: uvrA
A;Map position: 92 min
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; P-loop
F;31-38/Region: nucleotide-binding motif A (P-loop)
F;623-907/Domain: ATP-binding cassette homology <ABCE>
F;640-647/Region: nucleotide-binding motif A (P-loop)

Query Match 28.7%; Score 50; DB 1; Length 940;
Best Local Similarity 39.5%; Pred. No. 39;
Matches 15; Conservative 0; Mismatches 5; Indels 18; Gaps 2;

Qy 11 RTPDINPAWYTG----------------RGIRPVGRF 31
||| ||| ||| || | |||
Db 695 RTPRSNPATYTGVFTPVRELFAGVPESRARGYTP-GRF 731

RESULT 15
D71645
excinuclease ABC chain A (uvrA) RP835 - Rickettsia prowazekii
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C;Accession: D71645
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499
A;Accession: D71645
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-953 <AND>
A;Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CAA15260.1; PID:g386136
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: uvrA; RP835
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; hydrolase
F;33-40/Region: nucleotide-binding motif A (P-loop)
F;635-919/Domain: ATP-binding cassette homology <ABCE>
F;652-659/Region: nucleotide-binding motif A (P-loop)

Query Match 28.7%; Score 50; DB 2; Length 953;
Best Local Similarity 39.5%; Pred. No. 40;
Matches 15; Conservative 1; Mismatches 4; Indels 18; Gaps 2;

Qy 11 RTPDINPAWYTG----------------RGIRPVGRF 31
||| ||| ||| || : ||||
Db 707 RTPRSNPATYTGAFTHIRDWFVELPESKARGYK-VGRF 743

Search completed: April 17, 2001, 15:45:54
Job time: 602 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:48:06 ; Search time 115.07 Seconds
(without alignments)
31.576 Million cell updates/sec

Title: US-09-446-543A-47
Perfect score: 174
Sequence: 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	107	61.5	117	13	Q9W624	Q9w624 carassius a
2	68	39.1	692	2	Q9L8J6	Q918j6 rhodospiril
3	61	35.1	1236	2	Q9JPA4	Q9jpa4 rhodocyclus
4	58	33.3	1292	2	Q9ZGE5	Q9zge5 heliobacill
5	56	32.2	428	6	O97859	O97859 bos taurus
6	53	30.5	54	4	Q9UJF9	Q9ujf9 homo sapien
7	53	30.5	223	2	O85605	O85605 prevotella
8	53	30.5	465	4	O60687	O60687 homo sapien
9	52	29.9	503	10	Q9LSC6	Q9lsc6 arabidopsis
10	52	29.9	940	2	Q9KUW5	Q9kuw5 vibrio chol
11	51.5	29.6	503	2	Q9KRY1	Q9kry1 vibrio chol
12	51	29.3	294	5	Q19530	Q19530 caenorhabdi
13	51	29.3	428	4	Q9UQ49	Q9uq49 homo sapien
14	51	29.3	629	4	Q9NQE1	Q9nqe1 homo sapien
15	50.5	29.0	176	3	Q08689	Q08689 saccharomyc
16	50.5	29.0	333	2	Q9RJI0	Q9rji0 streptomyce
17	50.5	29.0	414	2	O33480	O33480 propionibac
18	50.5	29.0	538	4	Q9Y4C9	Q9y4c9 homo sapien
19	50.5	29.0	548	10	Q9LFA0	Q9lfa0 arabidopsis
20	50	28.7	863	5	Q9VRV3	Q9vrv3 drosophila
21	50	28.7	965	2	Q9PAR9	Q9par9 xylella fas
22	49.5	28.4	306	2	Q9L4D4	Q914d4 xanthomonas
23	49	28.2	128	2	P74747	P74747 synechocyst
24	49	28.2	320	2	Q9L7X4	Q917x4 pseudomonas
25	49	28.2	419	4	Q9Y276	Q9y276 homo sapien
26	49	28.2	772	10	O81349	O81349 chlamydomon
27	49	28.2	1193	2	Q9Z5E0	Q9z5e0 rhodobacter
28	49	28.2	1193	2	Q9RFD5	Q9rfd5 rhodobacter
29	49	28.2	1330	2	Q55284	Q55284 synechocyst
30	49	28.2	1331	2	P73020	P73020 synechocyst
31	49	28.2	1379	10	Q07893	Q07893 antirrhinum
32	49	28.2	1380	10	Q40001	Q40001 hordeum vul
33	49	28.2	1381	10	Q39049	Q39049 arabidopsis
34	49	28.2	1382	10	O22435	O22435 nicotiana t
35	49	28.2	1383	10	O65808	O65808 glycine max
36	48.5	27.9	189	5	Q9VHU4	Q9vhu4 drosophila
37	48.5	27.9	430	10	Q9SK71	Q9sk71 arabidopsis
38	48	27.6	157	2	Q9JRC1	Q9jrc1 neisseria m
39	48	27.6	369	4	Q9UG93	Q9ug93 homo sapien
40	48	27.6	455	2	O33340	O33340 mycobacteri
41	48	27.6	498	10	Q9STP4	Q9stp4 arabidopsis
42	48	27.6	517	10	Q9STP3	Q9stp3 arabidopsis
43	48	27.6	770	4	Q9Y2S4	Q9y2s4 homo sapien
44	48	27.6	790	10	Q9M371	Q9m371 arabidopsis
45	48	27.6	798	4	O94961	O94961 homo sapien

ALIGNMENTS

RESULT 1
Q9W624
ID Q9W624 PRELIMINARY; PRT; 117 AA.
AC Q9W624;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE C-RF AMIDE PRECURSOR.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Satake H., Minakata H., Fujimoto M.;
RT "Carassius RFamide (C-RF amide).";
RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB020024; BAA76662.1; -.
SQ SEQUENCE 117 AA; 12879 MW; D5DC4CB22038C2B0 CRC64;

Query Match 61.5%; Score 107; DB 13; Length 117;
Best Local Similarity 57.7%; Pred. No. 4.5e-08;
Matches 15; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

```
Qy       6 HSMETRTPDINPAWYTGRGIRPVGRF 31
           |::: |:|:|:| || |||:||:|||
Db      50 HNVDNRSPEIDPFWYVGRGVRPIGRF 75
```

RESULT 2
Q9L8J6
ID Q9L8J6 PRELIMINARY; PRT; 692 AA.
AC Q9L8J6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BCHH (FRAGMENT).
GN BCHH.
OS Rhodospirillum rubrum.

```
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC   Rhodospirillum.
OX   NCBI_TaxID=1085;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R5;
RX   MEDLINE=20138142; PubMed=10671438;
RA   Cheng Y.S., Brantner C.A., Tsapin A., Collins M.L.P.;
RT   "Role of the H protein in assembly of the photochemical reaction
RT   center and intracytoplasmic membrane in Rhodospirillum rubrum.";
RL   J. Bacteriol. 182:1200-1207(2000).
DR   EMBL; AF202319; AAF37352.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   692 AA;  75453 MW;  96430AE93BF35680 CRC64;


  Query Match             39.1%;  Score 68;  DB 2;  Length 692;
  Best Local Similarity   43.8%;  Pred. No. 0.13;
  Matches   14;  Conservative    4;  Mismatches    8;  Indels    6;  Gaps    1;

Qy       3 AHQHSMETRTPDINPAWYTG------RGIRPV 28
           | | |:|||  :|| || |       |:| :
Db     568 AEQVSLETRTRALNPKWYEGMLAHGYEGVRQI 599


RESULT   3
Q9JPA4
ID   Q9JPA4      PRELIMINARY;      PRT;  1236 AA.
AC   Q9JPA4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   MG PROTOPORPHYRIN METHYL TRANSFERASE.
GN   BCHH.
OS   Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).
OC   Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC   Rubrivivax.
OX   NCBI_TaxID=28068;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL144;
RA   Nagashima K.V., Shimada K., Matsuura K.;
RT   "Phylogenetic analysis of photosynthetic genes of Rhodocyclus
RT   gelatinosus: Possibility of horizontal gene transfer in purple
RT   bacteria.";
RL   Photosyn. Res. 36:185-191(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL144;
RX   MEDLINE=94132007; PubMed=8300574;
RA   Nagashima K.V., Matsuura K., Ohyama S., Shimada K.;
RT   "Primary structure and transcription of genes encoding B870 and
RT   photosynthetic reaction center apoproteins from Rubrivivax
RT   gelatinosus.";
RL   J. Biol. Chem. 269:2477-2484(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL144;
RA   Igarashi N., Shimada K., Matsuura K., Nagashima K.V.;
RT   "PHOTOSYNTHETIC GENE CLUSTER IN PURPLE BACTERIUM, RUBRIVIVAX
RT   GELATINOSUS.";
RL   (in) Garab G. (eds.);
RL   Photosynthesis. Mechanisms and Effects IV:2889-2892;
RL   Kluwer Academic Publishing (1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL144;
RA   Menin L., Yoshida M., Jaquinod M., Nagashima K.V., Matsuura K.,
RA   Parot P., Vermeglio A.;
RT   "Dark aerobic growth conditions induce the synthesis of a high
RT   midpoint potential cytochrome c8 in the photosynthetic bacterium
RT   Rubrivivax gelatinosus.";
RL   Biochemistry 0:0-0(1999).
DR   EMBL; AB034704; BAA94057.1; -.
KW   Transferase.
SQ   SEQUENCE   1236 AA;  134729 MW;  84051C045638520C CRC64;


  Query Match             35.1%;  Score 61;  DB 2;  Length 1236;
  Best Local Similarity   37.5%;  Pred. No. 2.4;
  Matches   12;  Conservative    6;  Mismatches    8;  Indels    6;  Gaps    1;

Qy       3 AHQHSMETRTPDINPAWYTG------RGIRPV 28
           : | ::|||  :|| || |       |:| :
Db    1112 SEQVALETRTRMLNPKWYEGMLEHGYEGVRQI 1143


RESULT   4
Q9ZGE5
ID   Q9ZGE5      PRELIMINARY;      PRT;  1292 AA.
AC   Q9ZGE5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MG CHELATASE SUBUNIT H BCHH.
GN   BCHH.
OS   Heliobacillus mobilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Heliobacterium group; Heliobacillus.
OX   NCBI_TaxID=28064;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99061957; PubMed=9843979;
RA   Xiong J., Inoue K., Bauer C.E.;
RT   "Tracking molecular evolution of photosynthesis by characterization of
RT   a major photosynthesis gene cluster from Heliobacillus mobilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
DR   EMBL; AF080002; AAC84033.1; -.
SQ   SEQUENCE   1292 AA;  144853 MW;  323AA0517B07448D CRC64;


  Query Match             33.3%;  Score 58;  DB 2;  Length 1292;
  Best Local Similarity   42.3%;  Pred. No. 6.8;
  Matches   11;  Conservative    3;  Mismatches    6;  Indels    6;  Gaps    1;

Qy       9 ETRTPDINPAWYTG------RGIRPV 28
           ||||  :|| || |       |:| :
Db    1176 ETRTKTLNPKWYEGMLKHGYEGVREI 1201


RESULT   5
O97859
ID   O97859      PRELIMINARY;      PRT;   428 AA.
AC   O97859;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GANGLIOSIDE SIALIDASE.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99143165; PubMed=9988745;
RA   Miyagi T., Wada T., Iwamatsu A., Hata K., Yoshikawa Y., Tokuyama S.,
RA   Sawada M.;
RT   "Molecular cloning and characterization of a plasma membrane-
RT   associated sialidase specific for gangliosides.";
RL   J. Biol. Chem. 274:5004-5011(1999).
DR   EMBL; AB008184; BAA75071.1; -.
DR   INTERPRO; IPR002860; -.
DR   PFAM; PF02012; BNR; 3.
SQ   SEQUENCE   428 AA;  47916 MW;  418B34F3245A8F21 CRC64;
```

```
  Query Match             32.2%;  Score 56;  DB 6;  Length 428;
  Best Local Similarity   37.0%;  Pred. No. 4.1;
  Matches   10;  Conservative    6;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 RAHQHSMETRTPDINPAWYTGRGIRPV 28
           ||  ||:   : |:   |: || |:|:
Db     195 RARPHSLMIYSDDLGATWHHGRLIKPM 221


RESULT   6
Q9UJF9
ID   Q9UJF9      PRELIMINARY;      PRT;    54 AA.
AC   Q9UJF9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DJ479J7.3 (SUSHI-REPEAT PROTEIN (SRPUL)) (FRAGMENT).
GN   DJ479J7.3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lawlor S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035608; CAB55682.1; -.
FT   NON_TER      54     54
SQ   SEQUENCE   54 AA;  6110 MW;  E2F3C39F7B961A9F CRC64;


  Query Match             30.5%;  Score 53;  DB 4;  Length 54;
  Best Local Similarity   50.0%;  Pred. No. 1.2;
  Matches    8;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      12 TPDINPAWYTGRGIRP 27
           || : | || | |  |
Db      18 TPAVTPTWYAGSGYYP 33


RESULT   7
O85605
ID   O85605      PRELIMINARY;      PRT;   223 AA.
AC   O85605;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A (FRAGMENT).
GN   UVRA.
OS   Prevotella albensis.
OC   Bacteria; CFB group; Bacteroidaceae; Prevotella.
OX   NCBI_TaxID=77768;
RN   [1]
RP   SEQUENCE OF 9-191 FROM N.A.
RC   STRAIN=M384;
RA   Walker N.D., McEwan N.R., Wallace R.J.;
RT   "Prevotella albensis putative uvrA gene.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070990; AAC24131.2; -.
DR   INTERPRO; IPR001617; -.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Transport.
FT   NON_TER       1      1
FT   NON_TER     223    223
SQ   SEQUENCE   223 AA;  24579 MW;  2563BA0B0C2996C9 CRC64;


  Query Match             30.5%;  Score 53;  DB 2;  Length 223;
  Best Local Similarity   39.5%;  Pred. No. 5.5;
  Matches   15;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                 || :| |||
Db      56 RTPRSNPATYTGVFSDIRTLFVGLPEAKIRGYKP-GRF 92


RESULT   8
O60687
ID   O60687      PRELIMINARY;      PRT;   465 AA.
AC   O60687;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   SUSHI-REPEAT PROTEIN.
GN   SRPUL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kurosawa H., Inukai T., Inaba T., Goi K., Chang K.-S., Sinjyo T.,
RA   Rakestraw K.M., Naeve C.W., Look T.A.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF060567; AAC15765.1; -.
DR   INTERPRO; IPR000436; -.
DR   INTERPRO; IPR001128; -.
DR   PFAM; PF00084; sushi; 3.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
SQ   SEQUENCE   465 AA;  52971 MW;  4D752B187FF3EFB8 CRC64;


  Query Match             30.5%;  Score 53;  DB 4;  Length 465;
  Best Local Similarity   50.0%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      12 TPDINPAWYTGRGIRP 27
           || : | || | |  |
Db      18 TPAVTPTWYAGSGYYP 33


RESULT   9
Q9LSC6
ID   Q9LSC6      PRELIMINARY;      PRT;   503 AA.
AC   Q9LSC6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   EMB|CAB75482.1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT   features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:131-135(2000).
DR   EMBL; AB026649; BAB01091.1; -.
SQ   SEQUENCE   503 AA;  57928 MW;  F5E38CD1BA9C521A CRC64;


  Query Match             29.9%;  Score 52;  DB 10;  Length 503;
  Best Local Similarity   33.3%;  Pred. No. 18;
  Matches    9;  Conservative    8;  Mismatches   10;  Indels    0;  Gaps    0;
```

```
Qy      3 AHQHSMETRTPDINPAWYTGRGIRPVG 29
          ||||:: : : |   : : |  :||:|
Db    199 AHQHNLMSFSKDSEESNFCGACVRPIG 225


RESULT  10
Q9KUW5
ID   Q9KUW5      PRELIMINARY;      PRT;   940 AA.
AC   Q9KUW5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   EXCINUCLEASE ABC, SUBUNIT A.
GN   VC0394.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004127; AAF93567.1; -.
DR   TIGR; VC0394; -.
SQ   SEQUENCE   940 AA;  104327 MW;  84F93B9DF686F6F2 CRC64;


  Query Match             29.9%;  Score 52;  DB 2;  Length 940;
  Best Local Similarity   39.5%;  Pred. No. 36;
  Matches   15;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy     11 RTPDINPAWYTG----------------RGIRPVGRF 31
          |||  ||| |||                || :| |||
Db    694 RTPRSNPATYTGIFTPIRELFAGTQESRSRGYQP-GRF 730


RESULT  11
Q9KRY1
ID   Q9KRY1      PRELIMINARY;      PRT;   503 AA.
AC   Q9KRY1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   SUN/NUCLEOLAR PROTEIN FAMILY PROTEIN.
GN   VC1502.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004228; AAF94657.1; -.
DR   TIGR; VC1502; -.
SQ   SEQUENCE   503 AA;  55997 MW;  2ABD94A2356C9E48 CRC64;


  Query Match             29.6%;  Score 51.5;  DB 2;  Length 503;
  Best Local Similarity   36.7%;  Pred. No. 22;
  Matches   11;  Conservative    5;  Mismatches   13;  Indels    1;  Gaps    1;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGR 30
          | :  ||:|  |      |: || :|| |:
Db    435 SSSASHSVELDTTQAR-EWFMGRDVRPEGQ 463


RESULT  12
Q19530
ID   Q19530      PRELIMINARY;      PRT;   294 AA.
AC   Q19530; Q19606;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   F19H6.1 PROTEIN.
GN   F19H6.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mcmurray A.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z50873; CAA90762.1; -.
DR   EMBL; Z68115; CAA90762.1; JOINED.
DR   EMBL; Z68115; CAA92169.1; -.
DR   EMBL; Z50873; CAA92169.1; JOINED.
DR   HSSP; P24941; 1CKP.
DR   INTERPRO; IPR000719; -.
DR   INTERPRO; IPR001245; -.
DR   INTERPRO; IPR002290; -.
DR   PFAM; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
SQ   SEQUENCE   294 AA;  34006 MW;  071D46644D049F43 CRC64;


  Query Match             29.3%;  Score 51;  DB 5;  Length 294;
  Best Local Similarity   44.8%;  Pred. No. 14;
  Matches   13;  Conservative    2;  Mismatches   12;  Indels    2;  Gaps    1;

Qy      3 AHQHSMETRTPDINPA--WYTGRGIRPVG 29
          || ||      || ||  : || ||  :|
Db    139 AHMHSKRIMHRDIKPANVFITGNGIVKLG 167


RESULT  13
Q9UQ49
ID   Q9UQ49      PRELIMINARY;      PRT;   428 AA.
AC   Q9UQ49;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   GANGLIOSIDE SIALIDASE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Wada T., Yoshikawa Y., Tokuyama S., Kuwabara M., Akita H., Miyagi T.;
RT   "Cloning, expression, and chromosomal mapping of a human ganglioside
RT   sialidase.";
```

```
RL   Biochem. Biophys. Res. Commun. 261:21-27(1999).
DR   EMBL; AB008185; BAA82611.1; -.
DR   INTERPRO; IPR002860; -.
DR   PFAM; PF02012; BNR; 3.
SQ   SEQUENCE   428 AA;  48252 MW;  35D1DD9359A78C98 CRC64;


  Query Match             29.3%;  Score 51;  DB 4;  Length 428;
  Best Local Similarity   33.3%;  Pred. No. 21;
  Matches    9;  Conservative    6;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      2 RAHQHSMETRTPDINPAWYTGRGIRPV 28
          :   ||:   : |:   |: || |||:
Db    195 KTRPHSLMIYSDDLGVTWHHGRLIRPM 221


RESULT  14
Q9NQE1
ID   Q9NQE1      PRELIMINARY;      PRT;   629 AA.
AC   Q9NQE1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   NURAMINIDASE (EC 3.2.1.18).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Monti E., Bassi M.T., Papini N., Riboni M., Manzoni M., Venerando B.,
RA   Croci G., Preti A., Ballabio A., Tettamanti G., Borsani G.;
RT   "Identification and expression of NEU3, a novel human sialidase
RT   associated to the plasma membrane.";
RL   Biochem. J. 349:343-351(2000).
DR   EMBL; Y18563; CAB96131.1; -.
KW   Hydrolase; Glycosidase.
SQ   SEQUENCE   629 AA;  69702 MW;  97C5464B70E69B4B CRC64;


  Query Match             29.3%;  Score 51;  DB 4;  Length 629;
  Best Local Similarity   33.3%;  Pred. No. 32;
  Matches    9;  Conservative    6;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      2 RAHQHSMETRTPDINPAWYTGRGIRPV 28
          :   ||:   : |:   |: || |||:
Db    396 KTRPHSLMIYSDDLGVTWHHGRLIRPM 422


RESULT  15
Q08689
ID   Q08689      PRELIMINARY;      PRT;   176 AA.
AC   Q08689;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CHROMOSOME XV READING FRAME ORF YOR253W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97298311; PubMed=9153759;
RA   Jauniaux J.C., Poirey R.;
RT   "Sequencing analysis of a 36.8 kb fragment of yeast chromosome XV
RT   reveals 26 open reading frames including SEC63, CDC31, SUG2, GCD1,
RT   RBL2, PNT1, PAC1 and VPH1.";
RL   Yeast 13:483-487(1997).
DR   EMBL; Z75161; CAA99475.1; -.
DR   INTERPRO; IPR000182; -.
DR   INTERPRO; IPR000345; -.
DR   PFAM; PF00583; Acetyltransf; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
SQ   SEQUENCE   176 AA;  19727 MW;  4F09DC597A690BA0 CRC64;


  Query Match             29.0%;  Score 50.5;  DB 3;  Length 176;
  Best Local Similarity   31.2%;  Pred. No. 9.8;
  Matches   10;  Conservative    4;  Mismatches   15;  Indels    3;  Gaps    1;

Qy      1 SRAHQHSMETRTP---DINPAWYTGRGIRPVG 29
          |  |||::    |   |:   |:   |   ||
Db    120 SECHQHNVFVYLPAVDDLTKQWFIAHGFEQVG 151


Search completed: April 17, 2001, 15:48:07
Job time: 560 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:48:47 ; Search time 39.1 Seconds
(without alignments)
27.159 Million cell updates/sec

Title: US-09-446-543A-47
Perfect score: 174
Sequence: 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	174	100.0	83	1	PRRP_RAT	P81278 rattus norv
2	163	93.7	98	1	PRRP_BOVIN	P81264 bos taurus
3	149	85.6	87	1	PRRP_HUMAN	P81277 homo sapien
4	56	32.2	962	1	UVRA_METTH	O26543 methanobact
5	53	30.5	798	1	UNR_RAT	P18395 rattus norv
6	52	29.9	972	1	UVRA_MYCTU	P94972 mycobacteri
7	52	29.9	973	1	UVRA_RHIME	P56899 rhizobium m
8	51	29.3	943	1	UVRA_HAEIN	P44410 haemophilus
9	50	28.7	264	1	Y355_BUCAI	P57436 buchnera ap
10	50	28.7	482	1	R167_YEAST	P39743 saccharomyc
11	50	28.7	940	1	UVRA_ECOLI	P07671 escherichia
12	50	28.7	941	1	UVRA_SALTY	P37434 salmonella
13	50	28.7	953	1	UVRA_RICPR	Q9zcc3 rickettsia
14	49.5	28.4	374	1	YHHJ_ECOLI	P31993 escherichia
15	49	28.2	952	1	UVRA_THETH	Q56242 thermus aqu
16	49	28.2	970	1	UVRA_SYNY3	P73412 synechocyst
17	49	28.2	1014	1	UVRA_STRCO	Q9z507 streptomyce
18	48	27.6	569	1	UVRA_VITST	Q08518 vitreoscill
19	48	27.6	719	1	ARP_YEAST	P32770 saccharomyc
20	48	27.6	798	1	UNR_HUMAN	O75534 homo sapien
21	48	27.6	950	1	UVRA_NEIGO	Q50968 neisseria g
22	48	27.6	960	1	UVRA_TREPA	O83527 treponema p
23	48	27.6	1194	1	BCHH_RHOCA	P26162 rhodobacter
24	47.5	27.3	501	1	TRA2_MOUSE	P39429 mus musculu
25	47.5	27.3	941	1	GCSP_MYCTU	Q50601 mycobacteri
26	47	27.0	453	1	TBB2_GEOCN	P32925 geotrichum
27	47	27.0	926	1	UVRA_AQUAE	O66911 aquifex aeo
28	47	27.0	1083	1	T2D3_HUMAN	O00268 homo sapien
29	46.5	26.7	272	1	TRA2_DROVI	O02008 drosophila
30	46.5	26.7	652	1	TETP_CLOPE	Q46306 clostridium
31	46	26.4	322	1	GRP2_MOUSE	O89100 m grb2-rela
32	46	26.4	330	1	GRP2_HUMAN	O75791 h grb2-rela
33	46	26.4	925	1	UVRA_ZYMMO	O31151 zymomonas m
34	46	26.4	957	1	UVRA_BACSU	O34863 bacillus su
35	46	26.4	992	1	UVRA_MICLU	P13567 micrococcus
36	46	26.4	1016	1	UVRA_DEIRA	Q46577 deinococcus
37	45.5	26.1	486	1	DMPC_PSESP	P19059 pseudomonas
38	45	25.9	126	1	YBGS_ECOLI	P75758 escherichia
39	45	25.9	184	1	RL6_THEMA	Q9zae4 thermotoga
40	45	25.9	430	1	YAE9_SCHPO	Q09849 schizosacch
41	45	25.9	676	1	EXL1_HUMAN	Q92935 homo sapien
42	45	25.9	773	1	DPOL_THEGO	P56689 thermococcu
43	45	25.9	817	1	VRP1_YEAST	P37370 saccharomyc
44	45	25.9	1523	1	DPOL_THEFM	P74918 thermococcu
45	44.5	25.6	240	1	PLSC_HELPY	O25903 helicobacte

ALIGNMENTS

RESULT 1
PRRP_RAT
ID PRRP_RAT STANDARD; PRT; 83 AA.
AC P81278;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE RELEASING PEPTIDE PRRP20].
GN PRH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RL Nature 393:272-276(1998).
CC -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC ----------

CC ----------
DR EMBL; AB015418; BAA29026.1; -.
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 21 BY SIMILARITY.
FT PEPTIDE 22 52 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT PEPTIDE 33 52 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 52 52 AMIDATION (G-53 PROVIDE AMIDE GROUP).
SQ SEQUENCE 83 AA; 9215 MW; D0C75A264EEE4F29 CRC64;

Query Match 100.0%; Score 174; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 2e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
|||||||||||||||||||||||||||||||
Db 22 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 52

RESULT 2

```
PRRP_BOVIN
ID   PRRP_BOVIN     STANDARD;      PRT;    98 AA.
AC   P81264;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015417; BAA29025.1; -.
KW   Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     22
FT   PEPTIDE      23     53       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      33     53       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      53     53       AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   98 AA;  10544 MW;  08AC35A13B0FA908 CRC64;


  Query Match             93.7%;  Score 163;  DB 1;  Length 98;
  Best Local Similarity   93.5%;  Pred. No. 9.1e-17;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          ||||||||| |||||||||| ||||||||||
Db     23 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 53


RESULT   3
PRRP_HUMAN
ID   PRRP_HUMAN     STANDARD;      PRT;    87 AA.
AC   P81277;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015419; BAA29027.1; -.
DR   MIM; 602663; -.
KW   Hormone; Amidation; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   PEPTIDE      23     53       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      34     53       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      53     53       AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   87 AA;  9639 MW;  229A2F3F50CF981B CRC64;


  Query Match             85.6%;  Score 149;  DB 1;  Length 87;
  Best Local Similarity   83.9%;  Pred. No. 8.3e-15;
  Matches   26;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          || |:|||| ||||||||||  |||||||||
Db     23 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 53


RESULT   4
UVRA_METTH
ID   UVRA_METTH     STANDARD;      PRT;   962 AA.
AC   O26543;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR MTH443.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
```

```
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000828; AAB84949.1; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      38     45       ATP (POTENTIAL).
FT   NP_BIND     649    656       ATP (POTENTIAL).
FT   ZN_FING     748    774       C4-TYPE.
SQ   SEQUENCE   962 AA;  108395 MW;  2C0EF7FC41CCD060 CRC64;


  Query Match             32.2%;  Score 56;  DB 1;  Length 962;
  Best Local Similarity   42.1%;  Pred. No. 2.5;
  Matches   16;  Conservative    0;  Mismatches    4;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                ||  || |||
Db     703 RTPRSNPATYTGVFTHIRELFAQTPEARKRGYRP-GRF 739


RESULT   5
UNR_RAT
ID   UNR_RAT        STANDARD;      PRT;   798 AA.
AC   P18395;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UNR PROTEIN.
GN   UNR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=90370473; PubMed=2204029;
RA   Jeffers M., Paciucci R., Pellicer A.;
RT   "Characterization of unr; a gene closely linked to N-ras.";
RL   Nucleic Acids Res. 18:4891-4899(1990).
CC   -!- FUNCTION: RNA-BINDING PROTEIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC   ---------------------*-----------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52311; CAA36549.1; -.
DR   PIR; S11210; S11210.
DR   HSSP; P15277; 1MJC.
DR   InterPro; IPR002059; -.
DR   Pfam; PF00313; CSD; 8.
DR   PROSITE; PS00352; COLD_SHOCK; 4.
KW   RNA-binding; Repeat.
FT   DOMAIN       26     87       CSD 1.
FT   DOMAIN      136    179       CSD 2 (INCOMPLETE).
FT   DOMAIN      186    245       CSD 3.
FT   DOMAIN      297    337       CSD 4 (INCOMPLETE).
FT   DOMAIN      349    410       CSD 5.
FT   DOMAIN      447    507       CSD 6.
FT   DOMAIN      519    579       CSD 7.
FT   DOMAIN      610    670       CSD 8.
FT   DOMAIN      674    735       CSD 9.
SQ   SEQUENCE   798 AA;  88894 MW;  F484B3FA8B0995A4 CRC64;


  Query Match             30.5%;  Score 53;  DB 1;  Length 798;
  Best Local Similarity   43.5%;  Pred. No. 5.5;
  Matches   10;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       6 HSMETRTPDINPAWYTGRGIRPV 28
           ||:   | : ||  |:|: |||:
Db     583 HSVNGITEEANPTIYSGKVIRPL 605


RESULT   6
UVRA_MYCTU
ID   UVRA_MYCTU     STANDARD;      PRT;   972 AA.
AC   P94972;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR RV1638 OR MTCY06H11.02.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z85982; CAB06633.1; -.
DR   TubercuList; Rv1638; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      32     39       ATP (POTENTIAL).
FT   NP_BIND     654    661       ATP (POTENTIAL).
```

```
FT   ZN_FING     257    285       C4-TYPE (ATYPICAL).
FT   ZN_FING     753    779       C4-TYPE.
SQ   SEQUENCE   972 AA;  106131 MW;  8937A764E592D981 CRC64;


  Query Match             29.9%;  Score 52;  DB 1;  Length 972;
  Best Local Similarity   39.5%;  Pred. No. 9.4;
  Matches   15;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                ||  :| |||
Db     708 RTPRSNPATYTGVFDKIRTLFAATTEAKVRGYQP-GRF 744


RESULT   7
UVRA_RHIME
ID   UVRA_RHIME     STANDARD;      PRT;   973 AA.
AC   P56899;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RA   Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
RA   Boistard P., Gouzy J., Kahn D., Thebault P., Goffeau A.,
RA   Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
RA   Vandenbol M., Puehler A., Becker A., Weidner S.;
RL   Submitted (MAY-2000) to the SWISS-PROT data bank.
RN   [2]
RP   SEQUENCE OF 1-140 FROM N.A.
RC   STRAIN=2021;
RX   MEDLINE=99430868; PubMed=10503543;
RA   Tapias A., Barbe J.;
RT   "Regulation of divergent transcription from the uvrA-ssb promoters in
RT   Sinorhizobium meliloti.";
RL   Mol. Gen. Genet. 262:121-130(1999).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF125162; AAF03210.1; -.
DR   InterPro; IPR001617; -.
DR   PROSITE; PS00211; ABC_TRANSPORTER; PARTIAL.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      34     41       ATP (POTENTIAL).
FT   NP_BIND     662    669       ATP (POTENTIAL).
FT   ZN_FING     761    787       C4-TYPE.
FT   CONFLICT     19     19       G -> A (IN REF. 2).
FT   CONFLICT     67     67       F -> S (IN REF. 2).
SQ   SEQUENCE   973 AA;  107191 MW;  3E1A8B14527A47FE CRC64;


  Query Match             29.9%;  Score 52;  DB 1;  Length 973;
  Best Local Similarity   39.5%;  Pred. No. 9.4;
  Matches   15;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                ||  :| |||
Db     716 RTPRSNPATYTGAFTPIRDWFAGLPEAKARGYQP-GRF 752


RESULT   8
UVRA_HAEIN
ID   UVRA_HAEIN     STANDARD;      PRT;   943 AA.
AC   P44410; Q48151;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR HI0249.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NTHI N187;
RX   MEDLINE=97080495; PubMed=8921840;
RA   de la Morena M.L., Hendrixson D.R., St Geme J.W. III;
RT   "Isolation and characterization of the Haemophilus influenzae uvrA
RT   gene.";
RL   Gene 177:23-28(1996).
RN   [3]
RP   SEQUENCE OF 1-71 FROM N.A.
RC   STRAIN=NTHI TN106;
RX   MEDLINE=94341556; PubMed=8063092;
RA   Jarosik G.P., Hansen E.J.;
RT   "Cloning and sequencing of the Haemophilus influenzae ssb gene
RT   encoding single-strand DNA-binding protein.";
RL   Gene 146:101-103(1994).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U32711; AAC21915.1; -.
DR   EMBL; U33877; AAC44592.1; -.
DR   EMBL; U04997; AAA60462.1; -.
DR   TIGR; HI0249; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      31     38       ATP (POTENTIAL).
FT   NP_BIND     640    647       ATP (POTENTIAL).
FT   ZN_FING     253    280       C4-TYPE.
FT   ZN_FING     740    766       C4-TYPE.
FT   CONFLICT    163    163       V -> L (IN REF. 2).
FT   CONFLICT    236    236       E -> D (IN REF. 2).
FT   CONFLICT    425    425       R -> K (IN REF. 2).
FT   CONFLICT    463    463       I -> M (IN REF. 2).
FT   CONFLICT    514    514       E -> Q (IN REF. 2).
FT   CONFLICT    661    661       A -> T (IN REF. 2).
FT   CONFLICT    928    928       T -> E (IN REF. 2).
FT   CONFLICT    935    943       FLKPILEKP -> PLP (IN REF. 2).
SQ   SEQUENCE   943 AA;  104366 MW;  4DBA0DCFA602D465 CRC64;


  Query Match             29.3%;  Score 51;  DB 1;  Length 943;
  Best Local Similarity   39.5%;  Pred. No. 13;
  Matches   15;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                ||  | |||
Db     695 RTPRSNPATYTGLFTPIRELFAGVPEARARGYNP-GRF 731


RESULT   9
Y355_BUCAI
ID   Y355_BUCAI     STANDARD;      PRT;   264 AA.
AC   P57436;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PUTATIVE DEOXYRIBONUCLEASE BU355 (EC 3.1.21.-).
GN   BU355.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TOKYO 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- SIMILARITY: BELONGS TO THE TATD DNASE FAMILY. STRONG, TO E.COLI
CC       YCFH.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP001119; BAB13059.1; -.
DR   PROSITE; PS01137; TATD_1; 1.
DR   PROSITE; PS01090; TATD_2; 1.
DR   PROSITE; PS01091; TATD_3; FALSE_NEG.
KW   Hypothetical protein; Hydrolase; Nuclease.
SQ   SEQUENCE   264 AA;  30520 MW;  7F1DA900018E0AAC CRC64;


  Query Match             28.7%;  Score 50;  DB 1;  Length 264;
  Best Local Similarity   41.7%;  Pred. No. 4.7;
  Matches   10;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       8 METRTPDINPAWYTGRGIRPVGRF 31
           :|| :| ::|| | |:| :|   |
Db     210 IETDSPYLSPAPYRGKGNQPAYLF 233


RESULT  10
R167_YEAST
ID   R167_YEAST     STANDARD;      PRT;   482 AA.
AC   P39743;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   REDUCED VIABILITY UPON STARVATION PROTEIN 167.
GN   RVS167 OR YDR388W OR D9509.8.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X2180;
RX   MEDLINE=93330299; PubMed=8336735;
RA   Bauer F., Urdaci M., Aigle M., Crouzet M.;
RT   "Alteration of a yeast SH3 protein leads to conditional viability
RT   with defects in cytoskeletal and budding patterns.";
RL   Mol. Cell. Biol. 13:5070-5084(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E.,
RA   Berno A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA   Hunicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA   Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA   Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA   Winant A., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   ACTIN-BINDING.
RX   MEDLINE=95236199; PubMed=7719850;
RA   Amberg D.C., Basart E., Botstein D.;
RT   "Defining protein interactions with yeast actin in vivo.";
RL   Nat. Struct. Biol. 2:28-35(1995).
CC   -!- FUNCTION: COMPONENT OF A CYTOSKELETAL STRUCTURE THAT IS REQUIRED
CC       FOR THE FORMATION OF ENDOCYTIC VESICLES AT THE PLASMA MEMBRANE
CC       LEVEL. COULD BE IMPLICATED IN CYTOSKELETAL REORGANIZATION IN
CC       RESPONSE TO ENVIRONMENTAL STRESSES AND COULD ACT IN THE BUDDING
CC       SITE SELECTION MECHANISM. BINDS TO ACTIN.
CC   -!- SIMILARITY: TO YEAST RVS161.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92092; AAA35051.1; -.
DR   EMBL; U32274; AAB64830.1; -.
DR   PIR; S40887; S40887.
DR   HSSP; P04002; 1ATF.
DR   SGD; S0002796; RVS167.
DR   InterPro; IPR001452; -.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PROSITE; PS50002; SH3; 1.
KW   Cytoskeleton; SH3 domain; Transmembrane; Actin-binding.
FT   DOMAIN      292    427       ALA/GLY/PRO-RICH.
FT   TRANSMEM    344    367       POTENTIAL.
```

```
FT   DOMAIN      421    482       SH3.
SQ   SEQUENCE   482 AA;  52774 MW;  3F0AB53EBCC95A5B CRC64;


  Query Match             28.7%;  Score 50;  DB 1;  Length 482;
  Best Local Similarity   50.0%;  Pred. No. 8.8;
  Matches   11;  Conservative    3;  Mismatches    2;  Indels    6;  Gaps    2;

Qy      11 RTPDINPAWYTGR-----GIRP 27
           ||||:|  |:|||     |: |
Db     453 RTPDVN-EWWTGRYNGQQGVFP 473


RESULT  11
UVRA_ECOLI
ID   UVRA_ECOLI     STANDARD;      PRT;   940 AA.
AC   P07671; P76788;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR DINE.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86168204; PubMed=3007478;
RA   Husain I., van Houten B., Thomas D.C., Sancar A.;
RT   "Sequences of Escherichia coli uvrA gene and protein reveal two
RT   potential ATP binding sites.";
RL   J. Biol. Chem. 261:4895-4901(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94089392; PubMed=8265357;
RA   Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA   Daniels D.L.;
RT   "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT   region from 89.2 to 92.8 minutes.";
RL   Nucleic Acids Res. 21:5408-5417(1993).
RN   [3]
RP   SEQUENCE OF 1-25 FROM N.A.
RX   MEDLINE=83299251; PubMed=6310514;
RA   Backendorf C., Brandsma J.A., Kartasova T., van de Putte P.;
RT   "In vivo regulation of the uvrA gene: role of the '-10' and '-35'
RT   promoter regions.";
RL   Nucleic Acids Res. 11:5795-5810(1983).
RN   [4]
RP   SEQUENCE OF 1-14 FROM N.A.
RX   MEDLINE=82220077; PubMed=6283374;
RA   Sancar A., Sancar G.B., Rupp W.D., Little J.W., Mount D.W.;
RT   "LexA protein inhibits transcription of the E. coli uvrA gene in
RT   vitro.";
RL   Nature 298:96-98(1982).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=91208117; PubMed=1826851;
RA   Myles G.M., Sancar A.;
RT   "Isolation and characterization of functional domains of UvrA.";
RL   Biochemistry 30:3834-3840(1991).
RN   [6]
RP   MUTAGENESIS OF CYS-253.
RX   MEDLINE=89380205; PubMed=2550431;
RA   Navaratnam S., Myles G.M., Strange R.W., Sancar A.;
RT   "Evidence from extended X-ray absorption fine structure and site-
RT   specific mutagenesis for zinc fingers in UvrA protein of Escherichia
RT   coli.";
RL   J. Biol. Chem. 264:16067-16071(1989).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA.
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- MISCELLANEOUS: BINDS ABOUT 2 ZINC ATOMS/MOLECULE.
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M13495; AAA24754.1; -.
DR   EMBL; U00006; AAC43152.1; -.
DR   EMBL; AE000479; AAC77028.1; -.
DR   EMBL; X01621; CAA25764.1; -.
DR   EMBL; J01721; AAA24753.1; -.
DR   PIR; A23869; BVECUA.
DR   ECO2DBASE; H124.0; 6TH EDITION.
DR   EcoGene; EG11061; uvrA.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      31     38       ATP.
FT   NP_BIND     640    647       ATP.
FT   ZN_FING     253    280       C4-TYPE.
FT   ZN_FING     740    766       C4-TYPE.
FT   MUTAGEN     253    253       C->A,H,S: REDUCED ACTIVITY.
SQ   SEQUENCE   940 AA;  103867 MW;  D61AAEB6514B860C CRC64;


  Query Match             28.7%;  Score 50;  DB 1;  Length 940;
  Best Local Similarity   39.5%;  Pred. No. 18;
  Matches   15;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                ||  | |||
Db     695 RTPRSNPATYTGVFTPVRELFAGVPESRARGYTP-GRF 731


RESULT  12
UVRA_SALTY
ID   UVRA_SALTY     STANDARD;      PRT;   941 AA.
AC   P37434;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Alberti M., Li Y.F., Sancar A., Hearst J.E.;
RL   Submitted (AUG-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- MISCELLANEOUS: BINDS ABOUT 2 ZINC ATOMS/MOLECULE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M93014; AAA27250.1; -.
DR   StyGene; SG10413; uvrA.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      31     38       ATP.
FT   NP_BIND     640    647       ATP.
FT   ZN_FING     253    280       C4-TYPE.
FT   ZN_FING     740    766       C4-TYPE.
SQ   SEQUENCE   941 AA;  103928 MW;  C4AFC9F549060C26 CRC64;


  Query Match             28.7%;  Score 50;  DB 1;  Length 941;
  Best Local Similarity   39.5%;  Pred. No. 18;
  Matches   15;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                ||  | |||
Db     695 RTPRSNPATYTGVFTPVRELFAGVPESRSRGYTP-GRF 731


RESULT  13
UVRA_RICPR
ID   UVRA_RICPR     STANDARD;      PRT;   953 AA.
AC   Q9ZCC3;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR RP835.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ235273; CAA15260.1; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      33     40       ATP (POTENTIAL).
FT   NP_BIND     652    659       ATP (POTENTIAL).
FT   ZN_FING     752    778       C4-TYPE.
SQ   SEQUENCE   953 AA;  106287 MW;  6209A66241379421 CRC64;


  Query Match             28.7%;  Score 50;  DB 1;  Length 953;
  Best Local Similarity   39.5%;  Pred. No. 18;
  Matches   15;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                || : ||||
Db     707 RTPRSNPATYTGAFTHIRDWFVELPESKARGYK-VGRF 743


RESULT  14
YHHJ_ECOLI
ID   YHHJ_ECOLI     STANDARD;      PRT;   374 AA.
AC   P31993; P76703;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 41.1 KDA PROTEIN IN RHSB-PIT INTERGENIC REGION.
GN   YHHJ.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94316500; PubMed=8041620;
RA   Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT   "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT   region from 76.0 to 81.5 minutes.";
RL   Nucleic Acids Res. 22:2576-2586(1994).
RN   [2]
RP   SEQUENCE OF 225-374 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=93259920; PubMed=8387990;
RA   Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;
RT   "Rhs elements of Escherichia coli K-12: complex composites of shared
RT   and unique components that have different evolutionary histories.";
RL   J. Bacteriol. 175:2799-2808(1993).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ABC-2 SUBFAMILY OF INTEGRAL MEMBRANE
CC       PROTEINS.
CC   -!- SIMILARITY: TO E.COLI YBHR AND YBHS AND TO THE C-TERMINAL OF
CC       E.COLI YHIG.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U00039; AAB18460.1; ALT_INIT.
DR   EMBL; AE000424; AAC76510.1; ALT_INIT.
DR   EMBL; L02370; AAC61886.1; -.
DR   EcoGene; EG11767; yhhJ.
DR   InterPro; IPR000412; -.
DR   PROSITE; PS00890; ABC2_MEMBRANE; 1.
KW   Hypothetical protein; Transport; Transmembrane; Inner membrane.
FT   TRANSMEM     23     43       POTENTIAL.
```

```
FT   TRANSMEM    173    193       POTENTIAL.
FT   TRANSMEM    230    250       POTENTIAL.
FT   TRANSMEM    256    276       POTENTIAL.
FT   TRANSMEM    284    304       POTENTIAL.
FT   TRANSMEM    343    363       POTENTIAL.
SQ   SEQUENCE   374 AA;  41061 MW;  02895FB13F493391 CRC64;


  Query Match             28.4%;  Score 49.5;  DB 1;  Length 374;
  Best Local Similarity   47.4%;  Pred. No. 7.9;
  Matches    9;  Conservative    5;  Mismatches    2;  Indels    3;  Gaps    1;

Qy       7 SMETR---TPDINPAWYTG 22
           |:|||    |:::|||: |
Db     160 SLETRMRFNPNLDPAWFGG 178


RESULT  15
UVRA_THETH
ID   UVRA_THETH     STANDARD;      PRT;   952 AA.
AC   Q56242;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA.
OS   Thermus aquaticus (subsp. thermophilus).
OC   Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
OX   NCBI_TaxID=274;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HB8 / ATCC 27634;
RX   MEDLINE=96257202; PubMed=8675016;
RA   Yamamoto N., Kato R., Kuramitsu S.;
RT   "Cloning, sequencing and expression of the uvrA gene from an
RT   extremely thermophilic bacterium, Thermus thermophilus HB8.";
RL   Gene 171:103-106(1996).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D49911; BAA08652.1; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      31     38       ATP (POTENTIAL).
FT   NP_BIND     643    650       ATP (POTENTIAL).
FT   ZN_FING     253    280       C4-TYPE.
FT   ZN_FING     742    768       C4-TYPE.
SQ   SEQUENCE   952 AA;  105232 MW;  FBE2338B45CA96DD CRC64;


  Query Match             28.2%;  Score 49;  DB 1;  Length 952;
  Best Local Similarity   39.5%;  Pred. No. 25;
  Matches   15;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy      11 RTPDINPAWYTG----------------RGIRPVGRF 31
           |||  ||| |||                ||  | |||
Db     697 RTPRSNPATYTGVFDEIRDLFAKTPEARKRGYGP-GRF 733


Search completed: April 17, 2001, 15:48:48
Job time: 536 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:39:47 ; Search time 61.54 Seconds
(without alignments)
9.677 Million cell updates/sec

Title: US-09-446-543A-47
Perfect score: 174
Sequence: 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	174	100.0	31	3	US-09-105-678A-8	Sequence 8, Appli
2	174	100.0	31	3	US-09-105-678A-37	Sequence 37, Appl
3	174	100.0	31	4	US-09-172-353-4	Sequence 4, Appli
4	174	100.0	32	3	US-09-105-678A-38	Sequence 38, Appl
5	174	100.0	33	3	US-09-105-678A-39	Sequence 39, Appl
6	163	93.7	31	3	US-09-105-678A-7	Sequence 7, Appli
7	163	93.7	31	3	US-09-105-678A-31	Sequence 31, Appl
8	163	93.7	32	3	US-09-105-678A-32	Sequence 32, Appl
9	163	93.7	33	3	US-09-105-678A-33	Sequence 33, Appl
10	152	87.4	29	3	US-09-105-678A-29	Sequence 29, Appl
11	149	85.6	31	3	US-09-105-678A-9	Sequence 9, Appli
12	149	85.6	31	3	US-09-105-678A-43	Sequence 43, Appl
13	149	85.6	32	3	US-09-105-678A-44	Sequence 44, Appl
14	149	85.6	33	3	US-09-105-678A-45	Sequence 45, Appl
15	116	66.7	20	3	US-09-105-678A-40	Sequence 40, Appl
16	116	66.7	21	3	US-09-105-678A-41	Sequence 41, Appl
17	116	66.7	22	3	US-09-105-678A-42	Sequence 42, Appl
18	111	63.8	20	3	US-09-105-678A-34	Sequence 34, Appl
19	111	63.8	21	3	US-09-105-678A-35	Sequence 35, Appl
20	111	63.8	22	3	US-09-105-678A-36	Sequence 36, Appl
21	105	60.3	19	3	US-09-105-678A-30	Sequence 30, Appl
22	105	60.3	20	3	US-09-105-678A-46	Sequence 46, Appl
23	105	60.3	21	3	US-09-105-678A-47	Sequence 47, Appl
24	105	60.3	22	3	US-09-105-678A-48	Sequence 48, Appl
25	104	59.8	21	3	US-09-105-678A-28	Sequence 28, Appl
26	51	29.3	239	2	US-08-712-709-7	Sequence 7, Appli
27	51	29.3	239	2	US-09-049-671-3	Sequence 3, Appli
28	51	29.3	239	3	US-09-295-068-3	Sequence 3, Appli
29	51	29.3	239	3	US-09-111-444-7	Sequence 7, Appli
30	50	28.7	349	1	US-08-118-270-71	Sequence 71, Appl
31	50	28.7	349	5	PCT-US93-08528-71	Sequence 71, Appl
32	50	28.7	940	4	US-09-078-347A-1	Sequence 1, Appli
33	49	28.2	774	3	US-08-902-632-2	Sequence 2, Appli
34	49	28.2	774	3	US-09-073-354-1	Sequence 1, Appli
35	49	28.2	774	3	US-08-656-005A-1	Sequence 1, Appli
36	49	28.2	774	4	US-09-073-259-1	Sequence 1, Appli
37	49	28.2	774	4	US-09-363-095-1	Sequence 1, Appli
38	47.5	27.3	80	2	US-08-691-814B-31	Sequence 31, Appl
39	47.5	27.3	501	1	US-08-331-394-4	Sequence 4, Appli
40	47.5	27.3	501	1	US-08-250-858-4	Sequence 4, Appli
41	47.5	27.3	501	1	US-08-446-915-4	Sequence 4, Appli
42	47.5	27.3	501	2	US-08-744-139-4	Sequence 4, Appli
43	47.5	27.3	501	5	PCT-US95-06639-4	Sequence 4, Appli
44	46	26.4	330	2	US-08-815-176-1	Sequence 1, Appli
45	46	26.4	555	2	US-08-982-232-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-09-105-678A-8
; Sequence 8, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-8


  Query Match           100.0%;  Score 174;  DB 3;  Length 31;
  Best Local Similarity  100.0%;  Pred. No. 5.5e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT   2
US-09-105-678A-37
; Sequence 37, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-37


  Query Match             100.0%;  Score 174;  DB 3;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT   3
US-09-172-353-4
; Sequence 4, Application US/09172353
; Patent No. 6197530
; GENERAL INFORMATION:
;  APPLICANT: Stricker-Kongra, Alain
;  APPLICANT: Gu, Wei
;  TITLE OF INVENTION: GPR10 AS A TARGET FOR IDENTIFYING WEIGHT MODULATING COMPOUNDS
;  FILE REFERENCE: 07334/102001
;  CURRENT APPLICATION NUMBER: US/09/172,353
;  CURRENT FILING DATE: 1998-10-14
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 31
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-172-353-4


  Query Match             100.0%;  Score 174;  DB 4;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT   4
US-09-105-678A-38
; Sequence 38, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-38


  Query Match             100.0%;  Score 174;  DB 3;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT   5
US-09-105-678A-39
```

```
; Sequence 39, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-39


  Query Match             100.0%;  Score 174;  DB 3;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31


RESULT   6
US-09-105-678A-7
; Sequence 7, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-7


  Query Match             93.7%;  Score 163;  DB 3;  Length 31;
  Best Local Similarity   93.5%;  Pred. No. 2.6e-18;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          ||||||||| |||||||||| ||||||||||
Db      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31


RESULT   7
US-09-105-678A-31
; Sequence 31, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-31


  Query Match             93.7%;  Score 163;  DB 3;  Length 31;
  Best Local Similarity   93.5%;  Pred. No. 2.6e-18;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          ||||||||| |||||||||| ||||||||||
Db      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31


RESULT   8
US-09-105-678A-32
; Sequence 32, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-32


  Query Match             93.7%;  Score 163;  DB 3;  Length 32;
  Best Local Similarity   93.5%;  Pred. No. 2.7e-18;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          ||||||||| |||||||||| ||||||||||
Db      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31


RESULT   9
US-09-105-678A-33
; Sequence 33, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-33


  Query Match             93.7%;  Score 163;  DB 3;  Length 33;
  Best Local Similarity   93.5%;  Pred. No. 2.8e-18;
  Matches   29;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          ||||||||| |||||||||| ||||||||||
Db      1 SRAHQHSMEIRTPDINPAWYAGRGIRPVGRF 31


RESULT  10
US-09-105-678A-29
; Sequence 29, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-29


  Query Match             87.4%;  Score 152;  DB 3;  Length 29;
  Best Local Similarity   93.1%;  Pred. No. 1.1e-16;
  Matches   27;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMETRTPDINPAWYTGRGIRPVG 29
           ||||||||| |||||||||| ||||||||
Db       1 SRAHQHSMEIRTPDINPAWYAGRGIRPVG 29


RESULT  11
US-09-105-678A-9
; Sequence 9, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-9


  Query Match             85.6%;  Score 149;  DB 3;  Length 31;
  Best Local Similarity   83.9%;  Pred. No. 3.4e-16;
  Matches   26;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
           || |:|||| |||||||||| |||||||||
Db       1 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 31


RESULT  12
US-09-105-678A-43
; Sequence 43, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  43:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-43


  Query Match             85.6%;  Score 149;  DB 3;  Length 31;
  Best Local Similarity   83.9%;  Pred. No. 3.4e-16;
  Matches   26;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
           || |:|||| |||||||||| |||||||||
Db       1 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 31
```

```
RESULT  13
US-09-105-678A-44
; Sequence 44, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-44


  Query Match             85.6%;  Score 149;  DB 3;  Length 32;
  Best Local Similarity   83.9%;  Pred. No. 3.6e-16;
  Matches   26;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          || |:|||| |||||||||| ||||||||
Db      1 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 31


RESULT  14
US-09-105-678A-45
; Sequence 45, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-45


  Query Match             85.6%;  Score 149;  DB 3;  Length 33;
  Best Local Similarity   83.9%;  Pred. No. 3.7e-16;
  Matches   26;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          || |:|||| |||||||||| ||||||||
Db      1 SRTHRHSMEIRTPDINPAWYASRGIRPVGRF 31


RESULT  15
US-09-105-678A-40
; Sequence 40, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
```

; INFORMATION FOR SEQ ID NO: 40:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 20 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-09-105-678A-40

Query Match 66.7%; Score 116; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      12 TPDINPAWYTGRGIRPVGRF 31
           ||||||||||||||||||||
Db       1 TPDINPAWYTGRGIRPVGRF 20
```

Search completed: April 17, 2001, 15:39:48
Job time: 317 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:38:42 ; Search time 116.94 Seconds
(without alignments)
15.154 Million cell updates/sec

Title: US-09-446-543A-47
Perfect score: 174
Sequence: 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	174	100.0	31	18	W31384	Rat type G protein
2	174	100.0	31	20	W97233	Rat type ligand po
3	174	100.0	31	20	W87614	Rat 19P2 ligand.
4	174	100.0	31	20	W95173	Murine pituitary-d
5	174	100.0	31	20	W95174	Murine pituitary-d
6	174	100.0	31	21	B10355	Rat oxytocin secre
7	174	100.0	31	21	Y87504	Rat prolactin-rele
8	174	100.0	31	21	Y49292	19P2 ligand peptid
9	174	100.0	32	18	W31385	Rat type G protein
10	174	100.0	32	21	B10356	Rat oxytocin secre
11	174	100.0	33	18	W31386	Rat type G protein
12	174	100.0	33	21	B10357	Rat oxytocin secre
13	174	100.0	82	20	W95172	Murine pituitary-d
14	174	100.0	83	18	W31383	Rat type G protein
15	174	100.0	83	20	W97225	Rat type ligand po
16	174	100.0	83	21	B10354	Rat oxytocin secre
17	163	93.7	31	18	W31371	Bovine G protein-c
18	163	93.7	31	20	W97218	Bovine pituitary-d
19	163	93.7	31	20	W87613	Bovine 19P2 ligand
20	163	93.7	31	20	W95188	Bovine pituitary-d
21	163	93.7	31	21	B10347	Bovine oxytocin se
22	163	93.7	31	21	Y49290	19P2 ligand peptid
23	163	93.7	31	21	Y49298	19P2 ligand peptid
24	163	93.7	32	18	W31372	Bovine G protein-c
25	163	93.7	32	20	W95189	Bovine pituitary-d
26	163	93.7	32	21	B10348	Bovine oxytocin se
27	163	93.7	33	18	W31373	Bovine G protein-c
28	163	93.7	33	20	W95190	Bovine pituitary-d
29	163	93.7	33	21	B10349	Bovine oxytocin se
30	163	93.7	33	21	Y49297	19P2 ligand peptid
31	163	93.7	98	18	W31382	Bovine genome deri
32	163	93.7	98	18	W31368	Bovine G protein-c
33	163	93.7	98	20	W97224	Bovine genome-driv
34	163	93.7	98	20	W97217	Bovine pituitary-d
35	163	93.7	98	20	W95187	Bovine genome-deri
36	163	93.7	98	21	B10346	Bovine oxytocin se
37	163	93.7	98	21	B10353	Bovine oxytocin se
38	152	87.4	29	18	W31369	Bovine G protein-c
39	152	87.4	29	20	W95184	Bovine pituitary-d
40	149	85.6	31	18	W31391	Human type G prote
41	149	85.6	31	20	W97235	Human type ligand
42	149	85.6	31	20	W87615	Human 19P2 ligand.
43	149	85.6	31	21	B10362	Human oxytocin sec
44	149	85.6	31	21	Y49291	19P2 ligand peptid
45	149	85.6	32	18	W31392	Human type G prote

ALIGNMENTS

RESULT 1
W31384
ID W31384 standard; Peptide; 31 AA.
XX
AC W31384;
XX
DT 06-APR-1998 (first entry)
XX
DE Rat type G protein-coupled receptor ligand fragment 1.
XX
KW G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
OS Rat sp.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
DR N-PSDB; V02421.

XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2; Page 179; 258pp; English.
XX
CC This sequence represents a peptide fragment from a novel rat type
CC ligand polypeptide corresponding to amino acid residues 22 to 52 of the
CC sequence represented in W31383 and is used in an assay to monitor ligand
CC binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a prophylactic
CC or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC compounds which are capable of altering the binding activity of the
CC ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 174; DB 18; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.8e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmetrtpdinpawytgrgirpvgrf 31

RESULT 2
W97233
ID W97233 standard; peptide; 31 AA.
XX
AC W97233;
XX
DT 06-MAY-1999 (first entry)
XX
DE Rat type ligand polypeptide fragment.
XX
KW Rat type ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypoovarianism; gonecyst cacogenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmeniopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castilo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.
XX
OS Rattus sp.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI; 1999-105614/09.

XX
PT Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Claim 3; Page 153; 241pp; English.
XX
CC The present sequence represents a rat type ligand fragment. It
CC is used in the course of the invention. The specification describes
CC an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypoovarianism, gonecyst cacogenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can by used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmeniopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castilo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, irruption mole, abortion, unthrifty fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 174; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.8e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmetrtpdinpawytgrgirpvgrf 31

RESULT 3
W87614
ID W87614 standard; Peptide; 31 AA.
XX
AC W87614;
XX
DT 29-MAR-1999 (first entry)
XX
DE Rat 19P2 ligand.
XX
KW 19P2 ligand; G protein coupled receptor; pituitary;
KW prolactin releasing peptide; rat; dementia; breast cancer;
KW therapy.
XX
OS Rattus sp.
XX
PN EP887417-A2.
XX
PD 30-DEC-1998.
XX
PF 25-JUN-1998; 98EP-0111725.
XX
PR 27-JUN-1997; 97JP-0172118.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Moriya T, Nishimura O, Suenaga M, Tanaka Y;
XX
DR WPI; 1999-047884/05.
XX
PT Producing a 19P2 pituitary G protein receptor ligand - by cleavage
PT of a fusion protein, useful for preventing and treating dementia,
PT breast cancer, renal failure and autoimmune disease
XX
PS Claim 5; Page 34; 56pp; English.
XX

CC This is the amino acid sequence of the rat pituitary G
CC protein-coupled receptor ligand 19P2L. A method suitable for
CC commercial high-level production of 19P2L comprises expressing
CC the ligand in host cells as a recombinant fusion protein e.g. with
CC human basic fibroblast growth factor (see V83794-95) that has
CC been modified to include an N-terminal cysteine residue. The
CC ligand is released from the fusion by cyanylation followed by
CC ammonolysis. 19P2L has prolactin secretion-stimulating and (at
CC high doses) prolactin secretion-inhibiting properties. It can be
CC used in the treatment and prevention of various diseases including:
CC senile dementia, cerebrovascular dementia, and dementia associated
CC with: genealogical disorders (e.g. Alzheimer's disease, Parkinson's
CC disease, Pick's disease, Huntington's disease), infectious diseases
CC (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or
CC toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism,
CC intoxication by drugs, metal and organic compounds), tumourigenic
CC diseases (e.g. brain tumour), traumatic diseases (e.g. chronic
CC subarachnoidal heamorrhage, and other types of dementia, depression,
CC hyperactive child syndrome (microencephalopathy) and disturbance of
CC consciousness. It is also useful for prevention and treatment of
CC diseases associated with prolactin hypo and hypersecretion
CC respectively, including: hyperprolactinaemia, pituitary adenoma,
CC breast cancer, infertility, impotence and autoimmune disease
CC (hypersecretion disorders), and seminal vesicle hypoplasia,
CC osteoporosis, menopausal syndrome and renal failure (hyposecretion
CC disorders). The 19P2 polypeptide/amide is also useful as a test
CC reagent for study of the prolactin secretory function or as a
CC lactogogue in mammalian farm animals.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 174; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.8e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmetrtpdinpawytgrgirpvgrf 31

RESULT 4
W95173
ID W95173 standard; peptide; 31 AA.
XX
AC W95173;
XX
DT 10-MAR-1999 (first entry)
XX
DE Murine pituitary-derived ligand mature polypeptide sequence.
XX
KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal.
XX
OS Mus sp.
XX
PN WO9849295-A1.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-JP01923.
XX
PR 28-APR-1997; 97JP-0109974.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fukusumi S, Hinuma S;
XX
DR WPI; 1999-009423/01.
XX
PT New polypeptide ligand for orphan G protein coupled receptors - used
PT for treating disorders of central nervous system, pituitary and
PT pancreas, and for drug screening
XX
PS Disclosure; Page 134; 206pp; English.
XX
CC This represents the matured murine pituitary-derived ligand polypeptide
CC sequence. The polypeptide is a ligand for the G-protein coupled orphan
CC receptor designated GPR10 (human) or UHR-1 (rat). Cells transformed with
CC a vector containing the ligand polypeptide encoding DNA are used to
CC produce a recombinant ligand polypeptide. The ligand polypeptide, and its
CC fragments, modulate function of the pituitary, central nervous system,
CC pancreas and other tissues and can be used to screen for agents that
CC modulate binding of the polypeptide to the receptor; to quantify the
CC amount of receptor in a sample and to raise antibodies. They may also be
CC used therapeutically, e.g. to treat senile dementia; Alzheimer's,
CC Parkinson's or Huntington's diseases; Creutzfeld-Jakob disease; poisoning
CC by heavy metals or drugs; diabetes; schizophrenia; disorders of growth
CC hormone secretion; cancer; rheumatoid arthritis, epilepsy and many
CC others, also to improve post-operative nutritional status and as
CC vasopressor. Transgenic animals carrying the ligand polypeptide encoding
CC DNA or its mutein are used to study the function of the polypeptide-
CC expressing genes, as models of disease, for drug screening and as source
CC of cell lines. The ligand polypeptide DNA is used as a source of probes
CC and primers; to identify related sequences; in receptor-binding assays;
CC for production of Ab and antisera; in drug development; for gene therapy
CC and to develop transgenic animals.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 174; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.8e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmetrtpdinpawytgrgirpvgrf 31

RESULT 5
W95174
ID W95174 standard; Protein; 31 AA.
XX
AC W95174;
XX
DT 10-MAR-1999 (first entry)
XX
DE Murine pituitary-derived ligand polypeptide antigenic epitope.
XX
KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal; epitope.
XX
OS Mus sp.
XX
PN WO9849295-A1.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-JP01923.
XX
PR 28-APR-1997; 97JP-0109974.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fukusumi S, Hinuma S;

XX
DR WPI; 1999-009423/01.
XX
PT New polypeptide ligand for orphan G protein coupled receptors - used
PT for treating disorders of central nervous system, pituitary and
PT pancreas, and for drug screening
XX
PS Disclosure; Page 26; 206pp; English.
XX
CC The invention relates to a murine pituitary-derived ligand polypeptide
CC which is a ligand for the G-protein coupled orphan receptor designated
CC GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC the ligand polypeptide encoding DNA are used to produce a recombinant
CC ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC function of the pituitary, central nervous system, pancreas and other
CC tissues and can be used to screen for agents that modulate binding of the
CC polypeptide to the receptor; to quantify the amount of receptor in a
CC sample and to raise antibodies. They may also be used therapeutically,
CC e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC rheumatoid arthritis, epilepsy and many others, also to improve post-
CC operative nutritional status and as vasopressor. Transgenic animals
CC carrying the ligand polypeptide encoding DNA or its mutein are used to
CC study the function of the polypeptide-expressing genes, as models of
CC disease, for drug screening and as source of cell lines. The ligand
CC polypeptide DNA is used as a source of probes and primers; to identify
CC related sequences; in receptor-binding assays; for production of Ab and
CC antisera; in drug development; for gene therapy and to develop transgenic
CC animals. Sequences W95174 to W95178 represent antigenic epiotpes which
CC can be used for the preparation of anti-ligand polypeptide antibody.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 174; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.8e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmetrtpdinpawytgrgirpvgrf 31

RESULT 6
B10355
ID B10355 standard; peptide; 31 AA.
XX
AC B10355;
XX
DT 24-NOV-2000 (first entry)
XX
DE Rat oxytocin secretion promoting peptide SEQ ID NO: 18.
XX
KW Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX
OS Rattus sp.
XX
PN WO200038704-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-JP07199.
XX
PR 25-DEC-1998; 98JP-0369585.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-452298/39.
XX
PT Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine -
XX
PS Claim 3; Page 57; 72pp; Japanese.
XX
CC This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a rat peptide which acts as an oxytocin secretion
CC promoter.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 174; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.8e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 1 srahqhsmetrtpdinpawytgrgirpvgrf 31

RESULT 7
Y87504
ID Y87504 standard; protein; 31 AA.
XX
AC Y87504;
XX
DT 18-JUL-2000 (first entry)
XX
DE Rat prolactin-releasing peptide, PrRP.
XX
KW Prolactin-releasing peptide; PrRP; GPR10; G protein-coupled receptor;
KW feeding behaviour; food intake; modulation; antagonist; anorectic;
KW obesity; agonist; cachexia.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT Modified-site 31
FT /note= "C-terminal amide"
XX
PN WO200017641-A1.
XX
PD 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-US21243.
XX
PR 22-SEP-1998; 98US-0101380.
PR 14-OCT-1998; 98US-0172353.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Stricker-Kongrad A, Gu W;
XX
DR WPI; 2000-303231/26.
XX
PT Identifying modulators of body weight by a combination of a cell-free
PT or cell-based assay to identify modulators of GPR10, followed by an in
PT vivo assay for the compounds effect on e.g. feeding behavior -
XX
PS Example 2; Page 61; 82pp; English.
XX

```
CC   The invention relates to a method for identifying compounds useful for
CC   modulating body weight. The method comprises cell-free and/or cell-based
CC   assays that identify compounds which bind to and/or activate or inhibit
CC   the activity of GPR10, a G protein-coupled receptor. These assays are
CC   then followed by an in vivo assay of the effect of the compound on
CC   feeding behaviour, body weight or metabolic rate in a mammal. Prolactin-
CC   releasing peptide (PrRP; Y87504) is a ligand of GPR10. Binding of PrRP to
CC   GPR10 stimulates a signal transduction cascade, which results in an
CC   increase in food intake. Compounds identified using the method of the
CC   invention are useful for the modulation of body weight. Antagonists of
CC   GPR10 can be used to treat obesity, while GPR10 agonists can be used to
CC   treat cachexia. The present sequence represents rat PrRP.
XX
SQ   Sequence   31 AA;


  Query Match             100.0%;  Score 174;  DB 21;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmetrtpdinpawytgrgirpvgrf 31


RESULT   8
Y49292
ID   Y49292 standard; peptide; 31 AA.
XX
AC   Y49292;
XX
DT   22-FEB-2000  (first entry)
XX
DE   19P2 ligand peptide fragment.
XX
KW   Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW   pituitary; regulatory mechanism; central nervous system; pancreatic.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   31
FT                   /note= "C-terminal amide"
XX
PN   WO9960112-A1.
XX
PD   25-NOV-1999.
XX
PF   20-MAY-1999;   99WO-JP02650.
XX
PR   21-MAY-1998;   98JP-0140293.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-039381/03.
XX
PT   New monoclonal antibodies,  useful in diagnosis, as drugs and  in
PT   studying diseases related to ligand abnormality  -
XX
PS   Disclosure; Page 26; 73pp; Japanese.
XX
CC   The invention provides a monoclonal antibody which has a specific
CC   reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC   derivative. The antibodies can be used in diagnosis or to treat or
CC   prevent diseases associated with abnormality in the pituitary function
CC   regulatory mechanism (e.g. promotion of prolactin secretion), central
CC   nervous regulatory mechanism, and pancreatic function regulatory
CC   mechanism. The antibody-based immunoassay can also be applied in
CC   clarifying the physiological functions of the ligand and its derivative.
CC   Sequences Y49290-302 represent peptide fragments of the 19P2 ligand.
XX
SQ   Sequence   31 AA;


  Query Match             100.0%;  Score 174;  DB 21;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmetrtpdinpawytgrgirpvgrf 31


RESULT   9
W31385
ID   W31385 standard; Peptide; 32 AA.
XX
AC   W31385;
XX
DT   06-APR-1998  (first entry)
XX
DE   Rat type G protein-coupled receptor ligand fragment 2.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Rat sp.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02422.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 179; 258pp; English.
XX
CC   This sequence represents a peptide fragment from a novel rat type
CC   ligand polypeptide corresponding to amino acid residues 22 to 53 of the
CC   sequence represented in W31383 and is used in an assay to monitor ligand
CC   binding to the G protein-coupled receptor protein. Pharmaceutical
CC   compositions containing this ligand may be used as a pituitary function
CC   modulator, a central nervous system modulator or a pancreatic function
CC   modulator. This ligand could have specific applications as a prophylactic
CC   or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC   disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC   growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC   hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC   amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC   compounds which are capable of altering the binding activity of the
CC   ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ   Sequence   32 AA;
```

```
  Query Match             100.0%;  Score 174;  DB 18;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 6e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmetrtpdinpawytgrgirpvgrf 31


RESULT  10
B10356
ID   B10356 standard; peptide; 32 AA.
XX
AC   B10356;
XX
DT   24-NOV-2000  (first entry)
XX
DE   Rat oxytocin secretion promoting peptide SEQ ID NO: 19.
XX
KW   Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
KW   treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW   caesarean section; artificial fertilization; galactostasis; goat; pig;
KW   veterinary medicine; milk production.
XX
OS   Rattus sp.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine  -
XX
PS   Disclosure; Page 57; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a rat peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   32 AA;


  Query Match             100.0%;  Score 174;  DB 21;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 6e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmetrtpdinpawytgrgirpvgrf 31


RESULT  11
W31386
ID   W31386 standard; Peptide; 33 AA.
XX
AC   W31386;
XX
DT   06-APR-1998  (first entry)
XX
DE   Rat type G protein-coupled receptor ligand fragment 3.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Rat sp.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02423.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 179-180; 258pp; English.
XX
CC   This sequence represents a peptide fragment from a novel rat type
CC   ligand polypeptide corresponding to amino acid residues 22 to 54 of the
CC   sequence represented in W31383 and is used in an assay to monitor ligand
CC   binding to the G protein-coupled receptor protein. Pharmaceutical
CC   compositions containing this ligand may be used as a pituitary function
CC   modulator, a central nervous system modulator or a pancreatic function
CC   modulator. This ligand could have specific applications as a prophylactic
CC   or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC   disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC   growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC   hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC   amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC   compounds which are capable of altering the binding activity of the
CC   ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ   Sequence   33 AA;


  Query Match             100.0%;  Score 174;  DB 18;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
          |||||||||||||||||||||||||||||||
Db      1 srahqhsmetrtpdinpawytgrgirpvgrf 31


RESULT  12
B10357
ID   B10357 standard; peptide; 33 AA.
XX
```

```
AC   B10357;
XX
DT   24-NOV-2000  (first entry)
XX
DE   Rat oxytocin secretion promoting peptide SEQ ID NO: 20.
XX
KW   Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
KW   treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW   caesarean section; artificial fertilization; galactostasis; goat; pig;
KW   veterinary medicine; milk production.
XX
OS   Rattus sp.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine -
XX
PS   Disclosure; Page 58; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a rat peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   33 AA;


  Query Match           100.0%;  Score 174;  DB 21;  Length 33;
  Best Local Similarity  100.0%;  Pred. No. 6.2e-19;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
           |||||||||||||||||||||||||||||||
Db       1 srahqhsmetrtpdinpawytgrgirpvgrf 31


RESULT  13
W95172
ID   W95172 standard; Protein; 82 AA.
XX
AC   W95172;
XX
DT   10-MAR-1999  (first entry)
XX
DE   Murine pituitary-derived ligand polypeptide.
XX
KW   Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW   GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW   tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW   Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW   secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW   gene therapy; transgenic animal.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1..20
FT                    /note= "signal peptide"
FT   Protein          21..82
FT                    /note= "mature protein"
XX
PN   WO9849295-A1.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-JP01923.
XX
PR   28-APR-1997;   97JP-0109974.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fukusumi S,  Hinuma S;
XX
DR   WPI; 1999-009423/01.
DR   N-PSDB; V81228, V81229.
XX
PT   New polypeptide ligand for orphan G protein coupled receptors - used
PT   for treating disorders of central nervous system, pituitary and
PT   pancreas, and for drug screening
XX
PS   Claim 1; Page 133; 206pp; English.
XX
CC   This represnts a murine pituitary-derived ligand polypeptide which is a
CC   ligand for the G-protein coupled orphan receptor designated GPR10 (human)
CC   or UHR-1 (rat). Cells transformed with a vector containing the ligand
CC   polypeptide encoding DNA are used to produce a recombinant ligand
CC   polypeptide. The ligand polypeptide, and its fragments, modulate function
CC   of the pituitary, central nervous system, pancreas and other tissues and
CC   can be used to screen for agents that modulate binding of the polypeptide
CC   to the receptor; to quantify the amount of receptor in a sample and to
CC   raise antibodies. They may also be used therapeutically, e.g. to treat
CC   senile dementia; Alzheimer's, Parkinson's or Huntington's diseases;
CC   Creutzfeld-Jakob disease; poisoning by heavy metals or drugs; diabetes;
CC   schizophrenia; disorders of growth hormone secretion; cancer; rheumatoid
CC   arthritis, epilepsy and many others, also to improve post-operative
CC   nutritional status and as vasopressor. Transgenic animals carrying the
CC   ligand polypeptide encoding DNA or its mutein are used to study the
CC   function of the polypeptide-expressing genes, as models of disease, for
CC   drug screening and as source of cell lines. The ligand polypeptide DNA is
CC   used as a source of probes and primers; to identify related sequences; in
CC   receptor-binding assays; for production of Ab and antisera; in drug
CC   development; for gene therapy and to develop transgenic animals.
XX
SQ   Sequence   82 AA;


  Query Match           100.0%;  Score 174;  DB 20;  Length 82;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-18;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
           |||||||||||||||||||||||||||||||
Db      21 srahqhsmetrtpdinpawytgrgirpvgrf 51


RESULT  14
W31383
ID   W31383 standard; Protein; 83 AA.
XX
AC   W31383;
XX
DT   06-APR-1998  (first entry)
XX
DE   Rat type G protein-coupled receptor ligand encoded by pRAV3.
```

XX
KW G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
OS Rat sp.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
DR N-PSDB; V02420.
XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 3; Page 178; 258pp; English.
XX
CC This sequence represents a novel rat type ligand polypeptide encoded by
CC pRAV3 which is used in an assay to monitor ligand binding to the G
CC protein-coupled receptor protein. Pharmaceutical compositions
CC containing this ligand may be used as a pituitary function modulator, a
CC central nervous system modulator or a pancreatic function modulator. This
CC ligand could have specific applications as a prophylactic or therapeutic
CC agent for dementia, depression, hyperkinetic syndrome, disturbance of
CC consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
CC secretory disease, hyper- and polyphagia, hypercholesterolaemia,
CC hyperglyceridaemia, hyperlipidaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia,
CC amylotrophic lateral sclerosis, acute myocardial infarction,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
CC can also be developed to screen compounds which are capable of altering
CC the binding activity of the ligand thus affecting activation of the G
CC protein-coupled receptor protein.
XX
SQ Sequence 83 AA;

Query Match 100.0%; Score 174; DB 18; Length 83;
Best Local Similarity 100.0%; Pred. No. 1.8e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 22 srahqhsmetrtpdinpawytgrgirpvgrf 52

RESULT 15
W97225
ID W97225 standard; Peptide; 83 AA.
XX
AC W97225;
XX
DT 06-MAY-1999 (first entry)
XX
DE Rat type ligand polypeptide.
XX
KW G protein-coupled receptor; GPCR; hypoovarianism; gonecyst cacogenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation; modulation;
KW pituitary adenomatosis; brain tumour; emmeniopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castilo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia; prolactin secretion;
KW rat type ligand.
XX
OS Rattus sp.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI; 1999-105614/09.
DR N-PSDB; X15525.
XX
PT Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Disclosure; Page 152; 241pp; English.
XX
CC The present sequence represents a rat type ligand polypeptide. The
CC specification describes an agent for modulating prolactin secretion
CC which comprises a ligand polypeptide or a salt, for a G protein-coupled
CC receptor (GPCR) protein. The agents for promoting prolactin secretion
CC can be used for treating or preventing hypoovarianism, gonecyst
CC cacogenesis, menopausal syndrome, euthyroid or hypometabolism. They
CC can by used for promoting lactation in a domestic mammal and as an
CC aphrodisiac. The agents for inhibiting prolactin secretion can be used
CC for treating or preventing pituitary adenomatosis, brain tumour,
CC emmeniopathy, autoimmune disease, prolactinoma, infertility, impotence,
CC amenorrhea, galactorrhea, acromegaly, Chiari-Frommel syndrome, Argonz-del
CC Castilo syndrome, Forbes-Albright syndrome, lymphoma, Sheehan syndrome
CC or dyszoospermia. The inhibitory agents can also be used as
CC contraceptives. The agents for modulating placental function can be used
CC for treating or preventing choriocarcinoma, hydatid mole, irruption mole,
CC abortion, unthrifty fetus, abnormal saccharometabolism, abnormal
CC lipidmetabolism or oxytocia.
XX
SQ Sequence 83 AA;

Query Match 100.0%; Score 174; DB 20; Length 83;
Best Local Similarity 100.0%; Pred. No. 1.8e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRAHQHSMETRTPDINPAWYTGRGIRPVGRF 31
||||||||||||||||||||||||||||||
Db 22 srahqhsmetrtpdinpawytgrgirpvgrf 52

Search completed: April 17, 2001, 15:38:42
Job time: 349 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:48:07 ; Search time 115.07 Seconds
(without alignments)
20.372 Million cell updates/sec

Title: US-09-446-543A-50
Perfect score: 116
Sequence: 1 TPDINPAWYTGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	90	77.6	117	13	Q9W624	Q9w624 carassius a
2	53	45.7	54	4	Q9UJF9	Q9ujf9 homo sapien
3	53	45.7	465	4	O60687	O60687 homo sapien
4	49	42.2	419	4	Q9Y276	Q9y276 homo sapien
5	48	41.4	223	2	O85605	O85605 prevotella
6	48	41.4	790	10	Q9M371	Q9m371 arabidopsis
7	47	40.5	430	1	O27142	O27142 methanobact
8	47	40.5	940	2	Q9KUW5	Q9kuw5 vibrio chol
9	46.5	40.1	333	2	Q9RJI0	Q9rji0 streptomyce
10	46	39.7	333	2	Q9PH76	Q9ph76 xylella fas
11	46	39.7	527	5	O76383	O76383 caenorhabdi
12	46	39.7	555	3	O00050	O00050 aspergillus
13	46	39.7	820	5	Q9NE93	Q9ne93 leishmania
14	45	38.8	267	5	Q9V3E5	Q9v3e5 drosophila
15	45	38.8	330	4	Q9NRB7	Q9nrb7 homo sapien
16	45	38.8	767	5	Q20170	Q20170 caenorhabdi
17	45	38.8	965	2	Q9PAR9	Q9par9 xylella fas
18	44	37.9	105	2	Q9PNE9	Q9pne9 campylobact
19	44	37.9	306	2	Q9L4D4	Q914d4 xanthomonas
20	44	37.9	320	2	Q9L7X4	Q917x4 pseudomonas
21	44	37.9	343	3	O74569	O74569 coprinus ci
22	44	37.9	398	10	Q9SJR5	Q9sjr5 arabidopsis
23	44	37.9	414	2	O33480	O33480 propionibac
24	44	37.9	418	2	Q9RYI3	Q9ryi3 deinococcus
25	44	37.9	449	5	O01498	O01498 caenorhabdi
26	44	37.9	540	10	Q9LGZ0	Q9lgz0 oryza sativ
27	44	37.9	548	10	Q9LFA0	Q9lfa0 arabidopsis
28	44	37.9	775	1	Q9P9K4	Q9p9k4 pyrococcus
29	43.5	37.5	506	2	Q9ZAA1	Q9zaa1 pseudomonas
30	43.5	37.5	1501	10	Q9SD86	Q9sd86 arabidopsis
31	43.5	37.5	1679	5	Q24301	Q24301 drosophila
32	43	37.1	152	10	Q39562	Q39562 chlamydomon
33	43	37.1	309	5	O17234	O17234 caenorhabdi
34	43	37.1	429	2	Q9L0X9	Q9l0x9 streptomyce
35	43	37.1	498	10	Q9STP4	Q9stp4 arabidopsis
36	43	37.1	503	2	Q9KRY1	Q9kry1 vibrio chol
37	43	37.1	517	10	Q9STP3	Q9stp3 arabidopsis
38	43	37.1	627	2	Q9L8G6	Q918g6 clostridium
39	43	37.1	630	4	Q9NXS8	Q9nxs8 homo sapien
40	43	37.1	948	2	Q9JUS4	Q9jus4 neisseria m
41	43	37.1	949	2	Q9JZP1	Q9jzp1 neisseria m
42	43	37.1	1292	2	Q9ZGE5	Q9zge5 heliobacill
43	42.5	36.6	243	2	Q9R6H5	Q9r6h5 agrobacteri
44	42.5	36.6	443	5	Q19879	Q19879 caenorhabdi
45	42.5	36.6	506	2	Q9KH02	Q9kh02 azotobacter

ALIGNMENTS

RESULT 1
Q9W624
ID Q9W624 PRELIMINARY; PRT; 117 AA.
AC Q9W624;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE C-RF AMIDE PRECURSOR.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Satake H., Minakata H., Fujimoto M.;
RT "Carassius RFamide (C-RF amide).";
RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB020024; BAA76662.1; -.
SQ SEQUENCE 117 AA; 12879 MW; D5DC4CB22038C2B0 CRC64;

Query Match 77.6%; Score 90; DB 13; Length 117;
Best Local Similarity 65.0%; Pred. No. 2.3e-06;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 TPDINPAWYTGRGIRPVGRF 20
          :|:|:| || |||:||:|||
Db     56 SPEIDPFWYVGRGVRPIGRF 75
```

RESULT 2
Q9UJF9
ID Q9UJF9 PRELIMINARY; PRT; 54 AA.
AC Q9UJF9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE DJ479J7.3 (SUSHI-REPEAT PROTEIN (SRPUL)) (FRAGMENT).
GN DJ479J7.3.
OS Homo sapiens (Human).

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lawlor S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035608; CAB55682.1; -.
FT   NON_TER      54     54
SQ   SEQUENCE   54 AA;  6110 MW;  E2F3C39F7B961A9F CRC64;


  Query Match             45.7%;  Score 53;  DB 4;  Length 54;
  Best Local Similarity   50.0%;  Pred. No. 0.41;
  Matches    8;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRP 16
          || : | || | |  |
Db     18 TPAVTPTWYAGSGYYP 33


RESULT   3
O60687
ID   O60687      PRELIMINARY;      PRT;   465 AA.
AC   O60687;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   SUSHI-REPEAT PROTEIN.
GN   SRPUL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kurosawa H., Inukai T., Inaba T., Goi K., Chang K.-S., Sinjyo T.,
RA   Rakestraw K.M., Naeve C.W., Look T.A.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF060567; AAC15765.1; -.
DR   INTERPRO; IPR000436; -.
DR   INTERPRO; IPR001128; -.
DR   PFAM; PF00084; sushi; 3.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
SQ   SEQUENCE   465 AA;  52971 MW;  4D752B187FF3EFB8 CRC64;


  Query Match             45.7%;  Score 53;  DB 4;  Length 465;
  Best Local Similarity   50.0%;  Pred. No. 4.3;
  Matches    8;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRP 16
          || : | || | |  |
Db     18 TPAVTPTWYAGSGYYP 33


RESULT   4
Q9Y276
ID   Q9Y276      PRELIMINARY;      PRT;   419 AA.
AC   Q9Y276;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   H-BCS1.
GN   BCS1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99097350; PubMed=9878253;
RA   Petruzzella V., Tiranti V., Fernandez P., Ianna P., Carrozzo R.,
RA   Zeviani M.;
RT   "Identification and characterization of human cDNAs specific to BCS1,
RT   PET112, SCO1, COX15, and COX11, five genes involved in the formation
RT   and function of the mitochondrial respiratory chain.";
RL   Genomics 54:494-504(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=96207227; PubMed=8619474;
RA   Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
RT   "A 'double adaptor' method for improved shotgun library
RT   construction.";
RL   Anal. Biochem. 236:107-113(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=97264341; PubMed=9110174;
RA   Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
RA   Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
RT   "Large-scale concatenation cDNA sequencing.";
RL   Genome Res. 7:353-358(1997).
DR   EMBL; AF026849; AAD08638.1; -.
DR   EMBL; AF038195; AAB97365.1; -.
DR   INTERPRO; IPR001939; -.
DR   PFAM; PF00004; AAA; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   419 AA;  47534 MW;  7F0F98BA62F2CBB8 CRC64;


  Query Match             42.2%;  Score 49;  DB 4;  Length 419;
  Best Local Similarity   80.0%;  Pred. No. 15;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 NPAWYTGRGI 14
          || ||| |||
Db    211 NPKWYTDRGI 220


RESULT   5
O85605
ID   O85605      PRELIMINARY;      PRT;   223 AA.
AC   O85605;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A (FRAGMENT).
GN   UVRA.
OS   Prevotella albensis.
OC   Bacteria; CFB group; Bacteroidaceae; Prevotella.
OX   NCBI_TaxID=77768;
RN   [1]
RP   SEQUENCE OF 9-191 FROM N.A.
RC   STRAIN=M384;
RA   Walker N.D., McEwan N.R., Wallace R.J.;
RT   "Prevotella albensis putative uvrA gene.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070990; AAC24131.2; -.
DR   INTERPRO; IPR001617; -.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Transport.
FT   NON_TER       1      1
FT   NON_TER     223    223
SQ   SEQUENCE   223 AA;  24579 MW;  2563BA0B0C2996C9 CRC64;


  Query Match             41.4%;  Score 48;  DB 2;  Length 223;
  Best Local Similarity   37.8%;  Pred. No. 11;
  Matches   14;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy      1 TPDINPAWYTG----------------RGIRPVGRF 20
          ||  ||| |||                 || :| |||
Db     57 TPRSNPATYTGVFSDIRTLFVGLPEAKIRGYKP-GRF 92
```

```
RESULT   6
Q9M371
ID   Q9M371      PRELIMINARY;      PRT;   790 AA.
AC   Q9M371;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 87.4 KDA PROTEIN.
GN   F15G16.60.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   De Haan M., Maarse A.C., Grivell L.A., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL132959; CAB71097.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   790 AA;  87375 MW;  B222724B75690F30 CRC64;

  Query Match             41.4%;  Score 48;  DB 10;  Length 790;
  Best Local Similarity   47.4%;  Pred. No. 44;
  Matches    9;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYTGRGIRPVGRF 20
           |  ||  |  ||::| ||:
Db     366 PPHNPRTYGSRGLQPHGRW 384


RESULT   7
O27142
ID   O27142      PRELIMINARY;      PRT;   430 AA.
AC   O27142;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CONSERVED PROTEIN.
GN   MTH1070.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanobacterium.
OX   NCBI_TaxID=2166;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000877; AAB85559.1; -.
DR   INTERPRO; IPR002510; -.
DR   PFAM; PF01523; PmbA_TldD; 1.
SQ   SEQUENCE   430 AA;  46062 MW;  C4FAE47C111749E6 CRC64;

  Query Match             40.5%;  Score 47;  DB 1;  Length 430;
  Best Local Similarity   61.5%;  Pred. No. 32;
  Matches    8;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3 DINPAWYTGRGIR 15
           |||| |  ||  |
Db     191 DINPEWVAGRACR 203


RESULT   8
Q9KUW5
ID   Q9KUW5      PRELIMINARY;      PRT;   940 AA.
AC   Q9KUW5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   EXCINUCLEASE ABC, SUBUNIT A.
GN   VC0394.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004127; AAF93567.1; -.
DR   TIGR; VC0394; -.
SQ   SEQUENCE   940 AA;  104327 MW;  84F93B9DF686F6F2 CRC64;

  Query Match             40.5%;  Score 47;  DB 2;  Length 940;
  Best Local Similarity   37.8%;  Pred. No. 75;
  Matches   14;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy       1 TPDINPAWYTG----------------RGIRPVGRF 20
           ||  ||| |||                  || :| |||
Db     695 TPRSNPATYTGIFTPIRELFAGTQESRSRGYQP-GRF 730


RESULT   9
Q9RJI0
ID   Q9RJI0      PRELIMINARY;      PRT;   333 AA.
AC   Q9RJI0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 36.3 KDA PROTEIN.
GN   SCF73.06C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=A3(2);
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL121746; CAB57411.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   333 AA;  36312 MW;  884FF2861837FD76 CRC64;


  Query Match             40.1%;  Score 46.5;  DB 2;  Length 333;
  Best Local Similarity   35.5%;  Pred. No. 29;
  Matches   11;  Conservative    2;  Mismatches    7;  Indels   11;  Gaps    1;

Qy       1 TPDINPAWYT----------GRGIRPVGRF 20
           ||   ||| |             |:|| ||:
Db      19 TPWEEPAWRTEAIGWIEARLAAHGLRPTGRW 49


RESULT  10
Q9PH76
ID   Q9PH76      PRELIMINARY;      PRT;   333 AA.
AC   Q9PH76;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYDROXYBENZOATE OCTAPRENYLTRANSFERASE.
GN   XF0068.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-157(2000).
DR   EMBL; AE003860; AAF82881.1; -.
DR   INTERPRO; IPR000537; -.
DR   PFAM; PF01040; UbiA; 1.
DR   PROSITE; PS00943; UBIA; UNKNOWN_1.
SQ   SEQUENCE   333 AA;  37931 MW;  ECF3F4716C962B95 CRC64;


  Query Match             39.7%;  Score 46;  DB 2;  Length 333;
  Best Local Similarity   53.3%;  Pred. No. 34;
  Matches    8;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 INPAWYTGRGIRPVG 18
           ::| |   || ||||
Db      54 LDPYWKLARGDRPVG 68


RESULT  11
O76383
ID   O76383      PRELIMINARY;      PRT;   527 AA.
AC   O76383;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   C24G6.6 PROTEIN.
GN   C24G6.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Greco T., Bradshaw H., Keppler D.;
RT   "The sequence of C. elegans cosmid C24G6.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF067936; AAC19213.1; -.
DR   INTERPRO; IPR000205; -.
DR   INTERPRO; IPR002937; -.
DR   PFAM; PF01593; Amino_oxidase; 1.
SQ   SEQUENCE   527 AA;  59805 MW;  9FBB1FB84437C5CB CRC64;


  Query Match             39.7%;  Score 46;  DB 5;  Length 527;
  Best Local Similarity   58.3%;  Pred. No. 56;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYTGRG 13
           |::  ||| |||
Db     370 PNVLSAWYAGRG 381


RESULT  12
O00050
ID   O00050      PRELIMINARY;      PRT;   555 AA.
AC   O00050;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   TRANSPOSASE.
```

```
OS   Aspergillus awamori.
OC   Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;
OC   anamorphic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=105351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UVK143F DERIVED FROM NRRL3112; TRANSPOSON=VADER;
RX   MEDLINE=97156905; PubMed=9003286;
RA   Nyyssonen E., Amutan M., Enfield L., Stubbs J., Dunn-Coleman N.S.;
RT   "The transposable element Tan1 of Aspergillus niger var. awamori, a
RT   new member of the Fot1 family.";
RL   Mol. Gen. Genet. 253:50-56(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UVK143F DERIVED FROM NRRL3112; TRANSPOSON=VADER;
RX   MEDLINE=96207472; PubMed=8625427;
RA   Amutan M., Nyyssonen E., Stubbs J., Diaz-Torres M.R., Dunn-Coleman N.;
RT   "Identification and cloning of a mobile transposon from Aspergillus
RT   niger var. awamori.";
RL   Curr. Genet. 29:468-473(1996).
DR   EMBL; U58946; AAC49623.1; -.
SQ   SEQUENCE   555 AA;  62978 MW;  70904D2EE009EA33 CRC64;


  Query Match             39.7%;  Score 46;  DB 3;  Length 555;
  Best Local Similarity   57.1%;  Pred. No. 60;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 AWYTGRGIRPVGRF 20
           ||| |: | |  ||
Db     237 AWYEGQSIPPTWRF 250


RESULT  13
Q9NE93
ID   Q9NE93      PRELIMINARY;      PRT;   820 AA.
AC   Q9NE93;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 86.4 KDA PROTEIN.
GN   L787.05.
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRIEDLIN;
RA   Robben J., Aert R., Volckaert G., Ivens A.C., Quail M.,
RA   Rajandream M.A., Barrell B.G.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRIEDLIN;
RA   Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA   Smith D.F.;
RT   "A physical map of the Leishmania major Friedlin genome.";
RL   Genome Res. 8:135-145(1998).
DR   EMBL; AL163492; CAB86689.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   820 AA;  86388 MW;  F4CB0D9FDE666817 CRC64;


  Query Match             39.7%;  Score 46;  DB 5;  Length 820;
  Best Local Similarity   47.1%;  Pred. No. 92;
  Matches    8;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYTGRGIRPVG 18
           |  :|| |:| | :|:|
Db     406 PFTSPALYSGNGSQPLG 422


RESULT  14
Q9V3E5
ID   Q9V3E5      PRELIMINARY;      PRT;   267 AA.
AC   Q9V3E5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG4999 PROTEIN.
GN   CG4999.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Page-McCaw A.W., Tsang G., Rubin G.M.;
RT   "Sequencing Drosophila cDNAs related to tetraspanins.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003553; AAF50354.1; -.
DR   EMBL; AF220044; AAF23828.1; -.
DR   FLYBASE; FBgn0035936; CG4999.
DR   INTERPRO; IPR000301; -.
DR   PFAM; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
SQ   SEQUENCE   267 AA;  29879 MW;  DE91A9B2459BB65E CRC64;


  Query Match             38.8%;  Score 45;  DB 5;  Length 267;
  Best Local Similarity   46.7%;  Pred. No. 38;
  Matches    7;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;
```

```
Qy       3 DINPAWYTGRGIRPV 17
           | :|||  |:| | :
Db     171 DASPAWVNGKGNRTI 185


RESULT  15
Q9NRB7
ID   Q9NRB7      PRELIMINARY;      PRT;   330 AA.
AC   Q9NRB7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ADAPTER PROTEIN GRID.
GN   GRID.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20281669; PubMed=10820259;
RA   Ellis J.H., Ashman C., Burden M.N., Kilpatrick K.E., Morse M.A.,
RA   Hamblin P.A.;
RT   "GRID: a novel Grb-2-related adapter protein that interacts with the
RT   activated T cell costimulatory receptor CD28.";
RL   J. Immunol. 164:5805-5814(2000).
DR   EMBL; AF236119; AAF60319.1; -.
SQ   SEQUENCE   330 AA;  37943 MW;  74F4C8D849B56D55 CRC64;


  Query Match             38.8%;  Score 45;  DB 4;  Length 330;
  Best Local Similarity   50.0%;  Pred. No. 48;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5 NPAWYTGRGIRPVGRF 20
           ||:|:|||    :| |
Db     305 NPSWWTGRLHNKLGFF 320


Search completed: April 17, 2001, 15:48:08
Job time: 561 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:45:54 ; Search time 70.08 Seconds
(without alignments)
19.613 Million cell updates/sec

Title: US-09-446-543A-50
Perfect score: 116
Sequence: 1 TPDINPAWYTGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	51	44.0	962	2	H69157	excinuclease ABC c
2	49	42.2	1670	2	S71551	DNA-directed DNA p
3	48.5	41.8	664	2	F83376	conserved hypothet
4	48	41.4	790	2	T47959	hypothetical prote
5	47	40.5	430	1	B69009	conserved hypothet
6	47	40.5	453	2	S18597	tubulin beta chain
7	47	40.5	719	2	S61046	ARP1 protein - yea
8	47	40.5	940	2	A82329	excinuclease ABC,
9	47	40.5	972	2	A70619	excinuclease ABC c
10	46.5	40.1	652	1	S41522	tetracycline resis
11	46	39.7	330	2	JE0376	Grb-2 related adap
12	46	39.7	333	2	H82852	hydroxybenzoate oc
13	46	39.7	527	2	T33175	hypothetical prote
14	46	39.7	943	2	D64057	excinuclease ABC c
15	45	38.8	482	1	S40887	RVS167 protein - y
16	45	38.8	767	2	T21969	hypothetical prote
17	45	38.8	798	2	S11210	probable unr prote
18	45	38.8	940	1	BVECUA	excinuclease ABC c
19	45	38.8	953	2	D71645	excinuclease ABC c
20	45	38.8	965	2	C82560	excinuclease ABC s
21	44	37.9	105	2	D81319	hypothetical prote
22	44	37.9	184	2	E72248	ribosomal protein
23	44	37.9	418	2	F75587	probable glycosylt
24	44	37.9	548	2	T47548	hypothetical prote
25	44	37.9	771	2	C75023	DNA polymerase I P
26	44	37.9	945	2	E83117	excinuclease ABC s
27	44	37.9	952	1	T46550	excinuclease ABC c
28	44	37.9	970	2	S77349	excinuclease ABC c
29	44	37.9	1014	2	T36031	excinuclease ABC c
30	44	37.9	1312	2	S68593	DNA-directed DNA p
31	43.5	37.5	506	2	H83396	probable aldehyde
32	43.5	37.5	506	2	F83142	probable aldehyde
33	43.5	37.5	1501	2	T45623	hypothetical prote
34	43.5	37.5	1680	2	A43434	furin (EC 3.4.21.7
35	43	37.1	220	2	C83292	probable glutathio
36	43	37.1	241	1	S64445	hypothetical prote
37	43	37.1	309	2	T32376	hypothetical prote
38	43	37.1	476	2	G64720	probable amino aci
39	43	37.1	498	2	T09021	beta-glucosidase h
40	43	37.1	503	2	A82193	Sun/nucleolar prot
41	43	37.1	517	2	T09022	beta-glucosidase h
42	43	37.1	948	2	B81883	excinuclease ABC c
43	43	37.1	949	2	A81138	excinuclease ABC c
44	43	37.1	960	2	A71315	excinuclease ABC c
45	43	37.1	1282	2	T30577	DNA topoisomerase

ALIGNMENTS

RESULT 1
H69157
excinuclease ABC chain A - Methanobacterium thermoautotrophicum (strain Delta H)
N;Alternate names: uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 12-Nov-1999
C;Accession: H69157
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: H69157
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-962 <MTH>
A;Cross-references: GB:AE000828; GB:AE000666; NID:g2621504; PIDN:AAB84949.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH443
A;Start codon: TTG
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; P-loop
F;38-45/Region: nucleotide-binding motif A (P-loop)
F;632-915/Domain: ATP-binding cassette homology <ABCE>
F;649-656/Region: nucleotide-binding motif A (P-loop)

Query Match 44.0%; Score 51; DB 2; Length 962;
Best Local Similarity 40.5%; Pred. No. 9.8;
Matches 15; Conservative 0; Mismatches 4; Indels 18; Gaps 2;

```
Qy      1 TPDINPAWYTG----------------RGIRPVGRF 20
          ||  ||| |||                 || || |||
Db    704 TPRSNPATYTGVFTHIRELFAQTPEARKRGYRP-GRF 739
```

RESULT 2
S71551
DNA-directed DNA polymerase (EC 2.7.7.7) KOD, intein containing precursor - Pyrococcu
N;Contains: DNA endonuclease (EC 3.1.-.-) PI-I; DNA endonuclease (EC 3.1.-.-) PI-II;
C;Species: Pyrococcus sp.
A;Variety: strain KOD1
C;Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 20-Apr-2000
C;Accession: S71551
R;Kakihara, H.; Takagi, M.; Imanaka, T.
submitted to the EMBL Data Library, March 1994
A;Description: Cloning and analysis of the DNA polymerase gene from a new hyperthermo
A;Reference number: S71551
A;Accession: S71551

```
A;Molecule type: DNA
A;Residues: 1-1670 <KAK>
A;Cross-references: EMBL:D29671
C;Function: <NUCL>
A;Description: as DNA-directed DNA polymerase, catalyzes the polymerization of DNA at th
A;Note: DNA-directed DNA polymerase KOD
C;Function: <EN1>
A;Description: as DNA endonuclease PI-I, catalyzes the hydrolysis of internal phosphodie
A;Note: DNA endonuclease PI-PspI
C;Function: <EN2>
A;Description: as DNA endonuclease PI-II, catalyzes the hydrolysis of internal phosphodi
A;Note: DNA endonuclease PI-PspII
C;Superfamily: DNA-directed DNA polymerase KOD
C;Keywords: endonuclease; hydrolase; nucleotidyltransferase; protein splicing
F;1-406,767-851,1388-1670/Product: DNA-directed DNA polymerase KOD #status predicted <MA
F;1-406/Domain: DNA-directed DNA polymerase KOD extein 1 #status predicted <XT1>
F;407-766/Product: DNA endonuclease PI-I (pol KOD intein 1) #status predicted <MAT2>
F;767-851/Domain: DNA-directed DNA polymerase KOD extein 2 #status predicted <XT2>
F;852-1387/Product: DNA endonuclease PI-II (pol KOD intein 2) #status predicted <MAT3>
F;1388-1670/Domain: DNA-directed DNA polymerase KOD extein 3 #status predicted <XT3>
F;406-767/Cross-link: peptide (Arg-Ser) #status predicted
F;851-1388/Cross-link: peptide (Asn-Ser) #status predicted


  Query Match             42.2%;  Score 49;  DB 2;  Length 1670;
  Best Local Similarity   44.4%;  Pred. No. 35;
  Matches    8;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 DINPAWYTGRGIRPVGRF 20
           :| ||| ||  :  | |:
Db     294 EITPAWETGENLERVARY 311


RESULT   3
F83376
conserved hypothetical protein PA2151 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83376
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337
A;Accession: F83376
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-664 <STO>
A;Cross-references: GB:AE004642; GB:AE004091; NID:g9948163; PIDN:AAG05539.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2151


  Query Match             41.8%;  Score 48.5;  DB 2;  Length 664;
  Best Local Similarity   62.5%;  Pred. No. 16;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       1 TPDINPAWYTGRGIRP 16
           |||||| |:  |  ||
Db     478 TPDINP-WFLQRSGRP 492


RESULT   4
T47959
hypothetical protein F15G16.60 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47959
R;De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quet
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z24480
A;Accession: T47959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-790 <DEH>
A;Cross-references: EMBL:AL132959
A;Experimental source: cultivar Columbia; BAC clone F15G16
C;Genetics:
A;Map position: 3
A;Introns: 39/1; 678/2; 698/3; 773/2
A;Note: F15G16.60


  Query Match             41.4%;  Score 48;  DB 2;  Length 790;
  Best Local Similarity   47.4%;  Pred. No. 23;
  Matches    9;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYTGRGIRPVGRF 20
           |  ||  |  ||::| ||:
Db     366 PPHNPRTYGSRGLQPHGRW 384


RESULT   5
B69009
conserved hypothetical protein MTH1070 - Methanobacterium thermoautotrophicum (strain
C;Species: Methanobacterium thermoautotrophicum
C;Date: 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change 23-Jul-1999
C;Accession: B69009
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: B69009
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-430 <MTH>
A;Cross-references: GB:AE000877; GB:AE000666; NID:g2622157; PIDN:AAB85559.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1070
C;Superfamily: conserved hypothetical protein MTH1070


  Query Match             40.5%;  Score 47;  DB 1;  Length 430;
  Best Local Similarity   61.5%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3 DINPAWYTGRGIR 15
           |||| |  ||  |
Db     191 DINPEWVAGRACR 203


RESULT   6
S18597
tubulin beta chain - yeast (Geotrichum candidum)
C;Species: Geotrichum candidum
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 30-Sep-1993
C;Accession: S18597
R;Gold, S.E.; Casale, W.L.; Keen, N.T.
Mol. Gen. Genet. 230, 104-112, 1991
A;Title: Characterization of two beta-tubulin genes from Geotrichum candidum.
A;Reference number: S18596; MUID:92079883
A;Accession: S18597
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-453 <GOL>
A;Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 422-
C;Superfamily: tubulin
C;Keywords: microtubule


  Query Match             40.5%;  Score 47;  DB 2;  Length 453;
```

```
  Best Local Similarity   70.0%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       8 WYTGRGIRPV 17
           |||| |: ||
Db     399 WYTGEGMEPV 408


RESULT   7
S61046
ARP1 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein D1478; protein YDL167c
C;Species: Saccharomyces cerevisiae
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
C;Accession: S61046; S31139; S67719
R;Pohl, T.M.
submitted to the EMBL Data Library, November 1995
A;Reference number: S61010
A;Accession: S61046
A;Molecule type: DNA
A;Residues: 1-719 <POH>
A;Cross-references: EMBL:Z67750; NID:g1061256; PIDN:CAA91579.1; PID:g1061272
R;Wehner, E.P.; Rao, E.; Brendel, M.
Mol. Gen. Genet. 237, 351-358, 1993
A;Title: Molecular structure and genetic regulation of SFA, a gene responsible for resis
A;Reference number: S31138; MUID:93247548
A;Accession: S31139
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-492,'N',494-719 <WEH>
A;Cross-references: EMBL:X68020; NID:g577609; PIDN:CAA48159.1; PID:g288590
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992
R;Pohl, T.M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67708
A;Accession: S67719
A;Molecule type: DNA
A;Residues: 1-719 <POW>
A;Cross-references: EMBL:Z74215; NID:g1431265; PIDN:CAA98741.1; PID:e253076; PID:g143126
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:NRP1; ARP1
A;Cross-references: MIPS:YDL167c; SGD:S0002326
A;Map position: 4L


  Query Match             40.5%;  Score 47;  DB 2;  Length 719;
  Best Local Similarity   58.3%;  Pred. No. 29;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 AWYTGRGIRPVG 18
           :|:|  |:||||
Db     244 SWFTQYGVRPVG 255


RESULT   8
A82329
excinuclease ABC, chain A VC0394 [imported] - Vibrio cholerae (group O1 strain N16961)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C;Accession: A82329
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: A82329
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-940 <HEI>
A;Cross-references: GB:AE004127; GB:AE003852; NID:g9654808; PIDN:AAF93567.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0394
A;Map position: 1
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology


  Query Match             40.5%;  Score 47;  DB 2;  Length 940;
  Best Local Similarity   37.8%;  Pred. No. 39;
  Matches   14;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy       1 TPDINPAWYTG----------------RGIRPVGRF 20
             ||  ||| |||                || :| |||
Db     695 TPRSNPATYTGIFTPIRELFAGTQESRSRGYQP-GRF 730


RESULT   9
A70619
excinuclease ABC chain A - Mycobacterium tuberculosis (strain H37RV)
N;Alternate names: uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70619
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: A70619
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-972 <COL>
A;Cross-references: GB:Z85982; GB:AL123456; NID:g3261718; PIDN:CAB06633.1; PID:g18389
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: uvrA
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; hydrolase
F;32-39/Region: nucleotide-binding motif A (P-loop)
F;637-920/Domain: ATP-binding cassette homology <ABCE>
F;654-661/Region: nucleotide-binding motif A (P-loop)


  Query Match             40.5%;  Score 47;  DB 2;  Length 972;
  Best Local Similarity   37.8%;  Pred. No. 40;
  Matches   14;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy       1 TPDINPAWYTG----------------RGIRPVGRF 20
             ||  ||| |||                || :| |||
Db     709 TPRSNPATYTGVFDKIRTLFAATTEAKVRGYQP-GRF 744


RESULT  10
S41522
tetracycline resistance protein tetB - Clostridium perfringens
C;Species: Clostridium perfringens
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S41522
R;Sloan, J.; McMurry, L.M.; Lyras, D.; Levy, S.B.; Rood, J.I.
Mol. Microbiol. 11, 403-415, 1994
A;Title: The Clostridium perfringens Tet P determinant comprises two overlapping gene
amily of tetracycline-resistance determinants.
A;Reference number: S41522; MUID:94224158
A;Accession: S41522
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-652 <SLO>
A;Cross-references: GB:L20800; NID:g456030; PIDN:AAA20117.1; PID:g456035
C;Superfamily: translation elongation factor G; translation elongation factor Tu homo
C;Keywords: antibiotic resistance; GTP binding; P-loop
F;5-132/Domain: translation elongation factor Tu homology <ETU>
```

```
F;11-18/Region: nucleotide-binding motif A (P-loop)
F;129-132/Region: GTP-binding NKXD motif


  Query Match             40.1%;  Score 46.5;  DB 1;  Length 652;
  Best Local Similarity   55.6%;  Pred. No. 32;
  Matches   10;  Conservative    1;  Mismatches    6;  Indels    1;  Gaps    1;

Qy       3 DINPAWYT-GRGIRPVGR 19
           |:|| | | |  | | ||
Db     445 DLNPFWATVGLEIEPAGR 462


RESULT  11
JE0376
Grb-2 related adaptor protein 2 - human
C;Species: Homo sapiens (man)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 21-Jul-2000
C;Accession: JE0376
R;Qiu, M.; Hua, S.; Agrawal, M.; Li, G.; Cai, J.; Chan, E.; Zhou, H.; Luo, Y.; Liu, M.
Biochem. Biophys. Res. Commun. 253, 443-447, 1998
A;Title: Molecular cloning and expression of human Grap-2, a novel leukocyte-specific SH
A;Reference number: JE0376; MUID:99097254
A;Accession: JE0376
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-330 <QIU>
A;Cross-references: GB:AF102694; NID:g3860192; PIDN:AAD04926.1; PID:g3860193


  Query Match             39.7%;  Score 46;  DB 2;  Length 330;
  Best Local Similarity   50.0%;  Pred. No. 19;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5 NPAWYTGRGIRPVGRF 20
           ||:|:|||    :| |
Db     305 NPSWWTGRLHNKLGLF 320


RESULT  12
H82852
hydroxybenzoate octaprenyltransferase XF0068 [imported] - Xylella fastidiosa (strain 9a5
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
C;Accession: H82852
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H82852
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-333 <SIM>
A;Cross-references: GB:AE003860; GB:AE003849; NID:g9104830; PIDN:AAF82881.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0068
C;Superfamily: 4-hydroxybenzoate octaprenyltransferase


  Query Match             39.7%;  Score 46;  DB 2;  Length 333;
  Best Local Similarity   53.3%;  Pred. No. 19;
  Matches    8;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 INPAWYTGRGIRPVG 18
           ::| |   || ||||
Db      54 LDPYWKLARGDRPVG 68


RESULT  13
T33175
hypothetical protein C24G6.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33175
R;Greco, T.; Bradshaw, H.; Keppler, D.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid C24G6.
A;Reference number: Z21298
A;Accession: T33175
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-527 <GRE>
A;Cross-references: EMBL:AF067936; PIDN:AAC19213.1; GSPDB:GN00023; CESP:C24G6.6
A;Experimental source: strain Bristol N2; clone C24G6
C;Genetics:
A;Gene: CESP:C24G6.6
A;Map position: 5
A;Introns: 20/3; 77/1; 129/2; 208/3; 470/2


  Query Match             39.7%;  Score 46;  DB 2;  Length 527;
  Best Local Similarity   58.3%;  Pred. No. 30;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYTGRG 13
           |::  ||| |||
Db     370 PNVLSAWYAGRG 381


RESULT  14
D64057
excinuclease ABC chain A - Haemophilus influenzae
N;Alternate names: uvrA protein
N;Contains: excision endonuclease ABC (EC 3.1.-.-) chain A
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 12-Nov-1999
C;Accession: D64057; JC5157
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: D64057
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-943 <TIGR>
A;Cross-references: GB:U32711; GB:L42023; NID:g1573214; PIDN:AAC21915.1; PID:g1573215
A;Experimental source: strain Rd KW20
R;de la Morena, M.L.; Hendrixson, D.R.; St Geme III, J.W.
Gene 177, 23-28, 1996
A;Title: Isolation and characterization of the Haemophilus influenzae uvrA gene.
A;Reference number: JC5157; MUID:97080495
A;Accession: JC5157
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-162,'L',164-235,'D',237-424,'K',426-462,'M',464-513,'Q',515-660,'T',662
A;Cross-references: GB:U33877; NID:g1144488; PIDN:AAC44592.1; PID:g1144489
A;Experimental source: strain N187
```

C;Genetics:
A;Gene: uvrA
C;Function:
A;Description: has ATPase and DNA binding activity; involved in DNA repair
C;Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C;Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; P-loop
F;31-38/Region: nucleotide-binding motif A (P-loop)
F;623-907/Domain: ATP-binding cassette homology <ABCE>
F;640-647/Region: nucleotide-binding motif A (P-loop)

Query Match 39.7%; Score 46; DB 2; Length 943;
Best Local Similarity 37.8%; Pred. No. 55;
Matches 14; Conservative 0; Mismatches 5; Indels 18; Gaps 2;

```
Qy       1 TPDINPAWYTG----------------RGIRPVGRF 20
           ||  ||| |||                ||  | |||
Db     696 TPRSNPATYTGLFTPIRELFAGVPEARARGYNP-GRF 731
```

RESULT 15
S40887
RVS167 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YDR388w
C;Species: Saccharomyces cerevisiae
C;Date: 31-Mar-1992 #sequence_revision 06-Feb-1995 #text_change 21-Jul-2000
C;Accession: S40887; S69672
R;Bauer, F.; Urdaci, M.; Aigle, M.; Crouzet, M.
Mol. Cell. Biol. 13, 5070-5084, 1993
A;Title: Alteration of a yeast SH3 protein leads to conditional viability with defects i
A;Reference number: S40887; MUID:93330299
A;Accession: S40887
A;Molecule type: DNA
A;Residues: 1-482 <BAU>
A;Cross-references: EMBL:M92092; NID:g172615; PIDN:AAA35051.1; PID:g172616
R;Dietrich, F.S.
submitted to the EMBL Data Library, July 1995
A;Description: The sequence of S. cerevisiae cosmids 9481, 9509, 9926, 9461, and lambda
A;Reference number: S69665
A;Accession: S69672
A;Molecule type: DNA
A;Residues: 1-482 <DIE>
A;Cross-references: EMBL:U32274; NID:g927313; PIDN:AAB64830.1; PID:g927321; GSPDB:GN000
C;Genetics:
A;Gene: SGD:RVS167; MIPS:YDR388w
A;Cross-references: SGD:S0002796; MIPS:YDR388w
A;Map position: 4R
C;Superfamily: RVS167 protein; RVS161 protein homology; SH3 homology
C;Keywords: transmembrane protein
F;4-270/Domain: RVS161 protein homology <RVS>
F;292-422/Region: alanine/glycine/proline-rich
F;428-477/Domain: SH3 homology <SH3>

Query Match 38.8%; Score 45; DB 1; Length 482;
Best Local Similarity 47.6%; Pred. No. 39;
Matches 10; Conservative 3; Mismatches 2; Indels 6; Gaps 2;

```
Qy       1 TPDINPAWYTGR-----GIRP 16
           |||:|  |:|||     |: |
Db     454 TPDVN-EWWTGRYNGQQGVFP 473
```

Search completed: April 17, 2001, 15:45:55
Job time: 603 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:38:42 ; Search time 116.94 Seconds
(without alignments)
9.777 Million cell updates/sec

Title: US-09-446-543A-50
Perfect score: 116
Sequence: 1 TPDINPAWYTGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	116	100.0	20	18	W31387	Rat type G protein
2	116	100.0	20	20	W97234	Rat type ligand po
3	116	100.0	20	20	W95175	Murine pituitary-d
4	116	100.0	20	21	B10358	Rat oxytocin secre
5	116	100.0	20	21	Y49302	19P2 ligand peptid
6	116	100.0	21	18	W31388	Rat type G protein
7	116	100.0	21	21	B10359	Rat oxytocin secre
8	116	100.0	22	18	W31389	Rat type G protein
9	116	100.0	22	21	B10360	Rat oxytocin secre
10	116	100.0	31	18	W31384	Rat type G protein
11	116	100.0	31	20	W97233	Rat type ligand po
12	116	100.0	31	20	W87614	Rat 19P2 ligand.
13	116	100.0	31	20	W95173	Murine pituitary-d
14	116	100.0	31	20	W95174	Murine pituitary-d
15	116	100.0	31	21	B10355	Rat oxytocin secre
16	116	100.0	31	21	Y87504	Rat prolactin-rele
17	116	100.0	31	21	Y49292	19P2 ligand peptid
18	116	100.0	32	18	W31385	Rat type G protein
19	116	100.0	32	21	B10356	Rat oxytocin secre
20	116	100.0	33	18	W31386	Rat type G protein
21	116	100.0	33	21	B10357	Rat oxytocin secre
22	116	100.0	82	20	W95172	Murine pituitary-d
23	116	100.0	83	18	W31383	Rat type G protein
24	116	100.0	83	20	W97225	Rat type ligand po
25	116	100.0	83	21	B10354	Rat oxytocin secre
26	111	95.7	20	18	W31374	Bovine G protein-c
27	111	95.7	20	20	W97232	Bovine pituitary-d
28	111	95.7	20	20	W95191	Bovine pituitary-d
29	111	95.7	20	21	B10350	Bovine oxytocin se
30	111	95.7	20	21	Y49301	19P2 ligand peptid
31	111	95.7	21	18	W31375	Bovine G protein-c
32	111	95.7	21	20	W95192	Bovine pituitary-d
33	111	95.7	21	21	B10351	Bovine oxytocin se
34	111	95.7	22	18	W31376	Bovine G protein-c
35	111	95.7	22	20	W95193	Bovine pituitary-d
36	111	95.7	22	21	B10352	Bovine oxytocin se
37	111	95.7	31	18	W31371	Bovine G protein-c
38	111	95.7	31	20	W97218	Bovine pituitary-d
39	111	95.7	31	20	W87613	Bovine 19P2 ligand
40	111	95.7	31	20	W95188	Bovine pituitary-d
41	111	95.7	31	21	B10347	Bovine oxytocin se
42	111	95.7	31	21	Y49290	19P2 ligand peptid
43	111	95.7	31	21	Y49298	19P2 ligand peptid
44	111	95.7	32	18	W31372	Bovine G protein-c
45	111	95.7	32	20	W95189	Bovine pituitary-d

ALIGNMENTS

RESULT 1
W31387
ID W31387 standard; Peptide; 20 AA.
XX
AC W31387;
XX
DT 06-APR-1998 (first entry)
XX
DE Rat type G protein-coupled receptor ligand fragment 4.
XX
KW G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
OS Rat sp.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
DR N-PSDB; V02424.

XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2; Page 180; 258pp; English.
XX
CC This sequence represents a peptide fragment from a novel rat type
CC ligand polypeptide corresponding to amino acid residues 33 to 52 of the
CC sequence represented in W31383 and is used in an assay to monitor ligand
CC binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a prophylactic
CC or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC compounds which are capable of altering the binding activity of the
CC ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 116; DB 18; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYTGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawytgrgirpvgrf 20

RESULT 2
W97234
ID W97234 standard; peptide; 20 AA.
XX
AC W97234;
XX
DT 06-MAY-1999 (first entry)
XX
DE Rat type ligand polypeptide fragment.
XX
KW Rat type ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypoovarianism; gonecyst cacogenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmeniopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castilo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.
XX
OS Rattus sp.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI; 1999-105614/09.

XX
PT Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Claim 3; Page 154; 241pp; English.
XX
CC The present sequence represents a rat type ligand fragment. It
CC is used in the course of the invention. The specification describes
CC an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypoovarianism, gonecyst cacogenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can by used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmeniopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castilo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, irruption mole, abortion, unthrifty fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 116; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYTGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawytgrgirpvgrf 20

RESULT 3
W95175
ID W95175 standard; Protein; 20 AA.
XX
AC W95175;
XX
DT 10-MAR-1999 (first entry)
XX
DE Murine pituitary-derived ligand polypeptide antigenic epitope.
XX
KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal; epitope.
XX
OS Mus sp.
XX
PN WO9849295-A1.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-JP01923.
XX
PR 28-APR-1997; 97JP-0109974.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fukusumi S, Hinuma S;
XX
DR WPI; 1999-009423/01.
XX
PT New polypeptide ligand for orphan G protein coupled receptors - used
PT for treating disorders of central nervous system, pituitary and

PT pancreas, and for drug screening
XX
PS Disclosure; Page 26; 206pp; English.
XX
CC The invention relates to a murine pituitary-derived ligand polypeptide
CC which is a ligand for the G-protein coupled orphan receptor designated
CC GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC the ligand polypeptide encoding DNA are used to produce a recombinant
CC ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC function of the pituitary, central nervous system, pancreas and other
CC tissues and can be used to screen for agents that modulate binding of the
CC polypeptide to the receptor; to quantify the amount of receptor in a
CC sample and to raise antibodies. They may also be used therapeutically,
CC e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC rheumatoid arthritis, epilepsy and many others, also to improve post-
CC operative nutritional status and as vasopressor. Transgenic animals
CC carrying the ligand polypeptide encoding DNA or its mutein are used to
CC study the function of the polypeptide-expressing genes, as models of
CC disease, for drug screening and as source of cell lines. The ligand
CC polypeptide DNA is used as a source of probes and primers; to identify
CC related sequences; in receptor-binding assays; for production of Ab and
CC antisera; in drug development; for gene therapy and to develop transgenic
CC animals. Sequences W95174 to W95178 represent antigenic epiotpes which
CC can be used for the preparation of anti-ligand polypeptide antibody.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 116; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYTGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawytgrgirpvgrf 20

RESULT 4
B10358
ID B10358 standard; peptide; 20 AA.
XX
AC B10358;
XX
DT 24-NOV-2000 (first entry)
XX
DE Rat oxytocin secretion promoting peptide SEQ ID NO: 21.
XX
KW Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX
OS Rattus sp.
XX
PN WO200038704-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-JP07199.
XX
PR 25-DEC-1998; 98JP-0369585.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-452298/39.
XX
PT Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine -
XX
PS Claim 5; Page 58; 72pp; Japanese.
XX
CC This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a rat peptide which acts as an oxytocin secretion
CC promoter.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 116; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYTGRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawytgrgirpvgrf 20

RESULT 5
Y49302
ID Y49302 standard; peptide; 20 AA.
XX
AC Y49302;
XX
DT 22-FEB-2000 (first entry)
XX
DE 19P2 ligand peptide fragment.
XX
KW Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW pituitary; regulatory mechanism; central nervous system; pancreatic.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT Modified-site 20
FT /note= "C-terminal amide"
XX
PN WO9960112-A1.
XX
PD 25-NOV-1999.
XX
PF 20-MAY-1999; 99WO-JP02650.
XX
PR 21-MAY-1998; 98JP-0140293.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-039381/03.
XX
PT New monoclonal antibodies, useful in diagnosis, as drugs and in
PT studying diseases related to ligand abnormality -
XX
PS Disclosure; Page 27; 73pp; Japanese.
XX
CC The invention provides a monoclonal antibody which has a specific
CC reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC derivative. The antibodies can be used in diagnosis or to treat or
CC prevent diseases associated with abnormality in the pituitary function
CC regulatory mechanism (e.g. promotion of prolactin secretion), central
CC nervous regulatory mechanism, and pancreatic function regulatory
CC mechanism. The antibody-based immunoassay can also be applied in
CC clarifying the physiological functions of the ligand and its derivative.

```
CC   Sequences Y49290-302 represent peptide fragments of the 19P2 ligand.
XX
SQ   Sequence   20 AA;


  Query Match             100.0%;  Score 116;  DB 21;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawytgrgirpvgrf 20


RESULT   6
W31388
ID   W31388 standard; Peptide; 21 AA.
XX
AC   W31388;
XX
DT   06-APR-1998  (first entry)
XX
DE   Rat type G protein-coupled receptor ligand fragment 5.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Rat sp.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02425.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 180; 258pp; English.
XX
CC   This sequence represents a peptide fragment from a novel rat type
CC   ligand polypeptide corresponding to amino acid residues 33 to 53 of the
CC   sequence represented in W31383 and is used in an assay to monitor ligand
CC   binding to the G protein-coupled receptor protein. Pharmaceutical
CC   compositions containing this ligand may be used as a pituitary function
CC   modulator, a central nervous system modulator or a pancreatic function
CC   modulator. This ligand could have specific applications as a prophylactic
CC   or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC   disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC   growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC   hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC   amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC   compounds which are capable of altering the binding activity of the
CC   ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ   Sequence   21 AA;


  Query Match             100.0%;  Score 116;  DB 18;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawytgrgirpvgrf 20


RESULT   7
B10359
ID   B10359 standard; peptide; 21 AA.
XX
AC   B10359;
XX
DT   24-NOV-2000  (first entry)
XX
DE   Rat oxytocin secretion promoting peptide SEQ ID NO: 22.
XX
KW   Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
KW   treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW   caesarean section; artificial fertilization; galactostasis; goat; pig;
KW   veterinary medicine; milk production.
XX
OS   Rattus sp.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine  -
XX
PS   Disclosure; Page 58; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a rat peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   21 AA;


  Query Match             100.0%;  Score 116;  DB 21;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawytgrgirpvgrf 20
```

```
RESULT   8
W31389
ID   W31389 standard; Peptide; 22 AA.
XX
AC   W31389;
XX
DT   06-APR-1998  (first entry)
XX
DE   Rat type G protein-coupled receptor ligand fragment 6.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Rat sp.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02426.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 181; 258pp; English.
XX
CC   This sequence represents a peptide fragment from a novel rat type
CC   ligand polypeptide corresponding to amino acid residues 33 to 54 of the
CC   sequence represented in W31383 and is used in an assay to monitor ligand
CC   binding to the G protein-coupled receptor protein. Pharmaceutical
CC   compositions containing this ligand may be used as a pituitary function
CC   modulator, a central nervous system modulator or a pancreatic function
CC   modulator. This ligand could have specific applications as a prophylactic
CC   or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC   disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC   growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC   hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC   amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC   compounds which are capable of altering the binding activity of the
CC   ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ   Sequence   22 AA;


  Query Match             100.0%;  Score 116;  DB 18;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           ||||||||||||||||||||
Db       1 tpdinpawytgrgirpvgrf 20


RESULT   9
B10360
ID   B10360 standard; peptide; 22 AA.
XX
AC   B10360;
XX
DT   24-NOV-2000  (first entry)
XX
DE   Rat oxytocin secretion promoting peptide SEQ ID NO: 23.
XX
KW   Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
KW   treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW   caesarean section; artificial fertilization; galactostasis; goat; pig;
KW   veterinary medicine; milk production.
XX
OS   Rattus sp.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine -
XX
PS   Disclosure; Page 59; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a rat peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   22 AA;


  Query Match             100.0%;  Score 116;  DB 21;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           ||||||||||||||||||||
Db       1 tpdinpawytgrgirpvgrf 20


RESULT   10
W31384
ID   W31384 standard; Peptide; 31 AA.
XX
AC   W31384;
XX
DT   06-APR-1998  (first entry)
XX
DE   Rat type G protein-coupled receptor ligand fragment 1.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
```

OS Rat sp.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
DR N-PSDB; V02421.
XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2; Page 179; 258pp; English.
XX
CC This sequence represents a peptide fragment from a novel rat type
CC ligand polypeptide corresponding to amino acid residues 22 to 52 of the
CC sequence represented in W31383 and is used in an assay to monitor ligand
CC binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a prophylactic
CC or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC compounds which are capable of altering the binding activity of the
CC ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 116; DB 18; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYTGRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawytgrgirpvgrf 31

RESULT 11
W97233
ID W97233 standard; peptide; 31 AA.
XX
AC W97233;
XX
DT 06-MAY-1999 (first entry)
XX
DE Rat type ligand polypeptide fragment.
XX
KW Rat type ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypoovarianism; gonecyst cacogenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmeniopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castilo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.
XX
OS Rattus sp.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI; 1999-105614/09.
XX
PT Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Claim 3; Page 153; 241pp; English.
XX
CC The present sequence represents a rat type ligand fragment. It
CC is used in the course of the invention. The specification describes
CC an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypoovarianism, gonecyst cacogenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can by used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmeniopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castilo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, irruption mole, abortion, unthrifty fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 116; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYTGRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawytgrgirpvgrf 31

RESULT 12
W87614
ID W87614 standard; Peptide; 31 AA.
XX
AC W87614;
XX
DT 29-MAR-1999 (first entry)
XX
DE Rat 19P2 ligand.
XX
KW 19P2 ligand; G protein coupled receptor; pituitary;
KW prolactin releasing peptide; rat; dementia; breast cancer;
KW therapy.
XX
OS Rattus sp.
XX

PN EP887417-A2.
XX
PD 30-DEC-1998.
XX
PF 25-JUN-1998; 98EP-0111725.
XX
PR 27-JUN-1997; 97JP-0172118.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Moriya T, Nishimura O, Suenaga M, Tanaka Y;
XX
DR WPI; 1999-047884/05.
XX
PT Producing a 19P2 pituitary G protein receptor ligand - by cleavage
PT of a fusion protein, useful for preventing and treating dementia,
PT breast cancer, renal failure and autoimmune disease
XX
PS Claim 5; Page 34; 56pp; English.
XX
CC This is the amino acid sequence of the rat pituitary G
CC protein-coupled receptor ligand 19P2L. A method suitable for
CC commercial high-level production of 19P2L comprises expressing
CC the ligand in host cells as a recombinant fusion protein e.g. with
CC human basic fibroblast growth factor (see V83794-95) that has
CC been modified to include an N-terminal cysteine residue. The
CC ligand is released from the fusion by cyanylation followed by
CC ammonolysis. 19P2L has prolactin secretion-stimulating and (at
CC high doses) prolactin secretion-inhibiting properties. It can be
CC used in the treatment and prevention of various diseases including:
CC senile dementia, cerebrovascular dementia, and dementia associated
CC with: genealogical disorders (e.g. Alzheimer's disease, Parkinson's
CC disease, Pick's disease, Huntington's disease), infectious diseases
CC (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or
CC toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism,
CC intoxication by drugs, metal and organic compounds), tumourigenic
CC diseases (e.g. brain tumour), traumatic diseases (e.g. chronic
CC subarachnoidal heamorrhage, and other types of dementia, depression,
CC hyperactive child syndrome (microencephalopathy) and disturbance of
CC consciousness. It is also useful for prevention and treatment of
CC diseases associated with prolactin hypo and hypersecretion
CC respectively, including: hyperprolactinaemia, pituitary adenoma,
CC breast cancer, infertility, impotence and autoimmune disease
CC (hypersecretion disorders), and seminal vesicle hypoplasia,
CC osteoporosis, menopausal syndrome and renal failure (hyposecretion
CC disorders). The 19P2 polypeptide/amide is also useful as a test
CC reagent for study of the prolactin secretory function or as a
CC lactogogue in mammalian farm animals.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 116; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYTGRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawytgrgirpvgrf 31

RESULT 13
W95173
ID W95173 standard; peptide; 31 AA.
XX
AC W95173;
XX
DT 10-MAR-1999 (first entry)
XX
DE Murine pituitary-derived ligand mature polypeptide sequence.
XX
KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal.
XX
OS Mus sp.
XX
PN WO9849295-A1.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-JP01923.
XX
PR 28-APR-1997; 97JP-0109974.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fukusumi S, Hinuma S;
XX
DR WPI; 1999-009423/01.
XX
PT New polypeptide ligand for orphan G protein coupled receptors - used
PT for treating disorders of central nervous system, pituitary and
PT pancreas, and for drug screening
XX
PS Disclosure; Page 134; 206pp; English.
XX
CC This represents the matured murine pituitary-derived ligand polypeptide
CC sequence. The polypeptide is a ligand for the G-protein coupled orphan
CC receptor designated GPR10 (human) or UHR-1 (rat). Cells transformed with
CC a vector containing the ligand polypeptide encoding DNA are used to
CC produce a recombinant ligand polypeptide. The ligand polypeptide, and its
CC fragments, modulate function of the pituitary, central nervous system,
CC pancreas and other tissues and can be used to screen for agents that
CC modulate binding of the polypeptide to the receptor; to quantify the
CC amount of receptor in a sample and to raise antibodies. They may also be
CC used therapeutically, e.g. to treat senile dementia; Alzheimer's,
CC Parkinson's or Huntington's diseases; Creutzfeld-Jakob disease; poisoning
CC by heavy metals or drugs; diabetes; schizophrenia; disorders of growth
CC hormone secretion; cancer; rheumatoid arthritis, epilepsy and many
CC others, also to improve post-operative nutritional status and as
CC vasopressor. Transgenic animals carrying the ligand polypeptide encoding
CC DNA or its mutein are used to study the function of the polypeptide-
CC expressing genes, as models of disease, for drug screening and as source
CC of cell lines. The ligand polypeptide DNA is used as a source of probes
CC and primers; to identify related sequences; in receptor-binding assays;
CC for production of Ab and antisera; in drug development; for gene therapy
CC and to develop transgenic animals.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 116; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYTGRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawytgrgirpvgrf 31

RESULT 14
W95174
ID W95174 standard; Protein; 31 AA.
XX
AC W95174;
XX
DT 10-MAR-1999 (first entry)
XX
DE Murine pituitary-derived ligand polypeptide antigenic epitope.
XX
KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;

KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal; epitope.
XX
OS Mus sp.
XX
PN WO9849295-A1.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-JP01923.
XX
PR 28-APR-1997; 97JP-0109974.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fukusumi S, Hinuma S;
XX
DR WPI; 1999-009423/01.
XX
PT New polypeptide ligand for orphan G protein coupled receptors - used
PT for treating disorders of central nervous system, pituitary and
PT pancreas, and for drug screening
XX
PS Disclosure; Page 26; 206pp; English.
XX
CC The invention relates to a murine pituitary-derived ligand polypeptide
CC which is a ligand for the G-protein coupled orphan receptor designated
CC GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
CC the ligand polypeptide encoding DNA are used to produce a recombinant
CC ligand polypeptide. The ligand polypeptide, and its fragments, modulate
CC function of the pituitary, central nervous system, pancreas and other
CC tissues and can be used to screen for agents that modulate binding of the
CC polypeptide to the receptor; to quantify the amount of receptor in a
CC sample and to raise antibodies. They may also be used therapeutically,
CC e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
CC diseases; Creutzfeld-Jakob disease; poisoning by heavy metals or drugs;
CC diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
CC rheumatoid arthritis, epilepsy and many others, also to improve post-
CC operative nutritional status and as vasopressor. Transgenic animals
CC carrying the ligand polypeptide encoding DNA or its mutein are used to
CC study the function of the polypeptide-expressing genes, as models of
CC disease, for drug screening and as source of cell lines. The ligand
CC polypeptide DNA is used as a source of probes and primers; to identify
CC related sequences; in receptor-binding assays; for production of Ab and
CC antisera; in drug development; for gene therapy and to develop transgenic
CC animals. Sequences W95174 to W95178 represent antigenic epiotpes which
CC can be used for the preparation of anti-ligand polypeptide antibody.
XX
SQ Sequence 31 AA;

```
  Query Match             100.0%;  Score 116;  DB 20;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           ||||||||||||||||||||
Db      12 tpdinpawytgrgirpvgrf 31
```

RESULT 15
B10355
ID B10355 standard; peptide; 31 AA.
XX
AC B10355;
XX
DT 24-NOV-2000 (first entry)
XX
DE Rat oxytocin secretion promoting peptide SEQ ID NO: 18.
XX
KW Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX
OS Rattus sp.
XX
PN WO200038704-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-JP07199.
XX
PR 25-DEC-1998; 98JP-0369585.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-452298/39.
XX
PT Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine -
XX
PS Claim 3; Page 57; 72pp; Japanese.
XX
CC This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a rat peptide which acts as an oxytocin secretion
CC promoter.
XX
SQ Sequence 31 AA;

```
  Query Match             100.0%;  Score 116;  DB 21;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           ||||||||||||||||||||
Db      12 tpdinpawytgrgirpvgrf 31
```

Search completed: April 17, 2001, 15:38:42
Job time: 349 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:39:48 ; Search time 61.54 Seconds
(without alignments)
6.243 Million cell updates/sec

Title: US-09-446-543A-50
Perfect score: 116
Sequence: 1 TPDINPAWYTGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	116	100.0	20	3	US-09-105-678A-40	Sequence 40, Appl
2	116	100.0	21	3	US-09-105-678A-41	Sequence 41, Appl
3	116	100.0	22	3	US-09-105-678A-42	Sequence 42, Appl
4	116	100.0	31	3	US-09-105-678A-8	Sequence 8, Appli
5	116	100.0	31	3	US-09-105-678A-37	Sequence 37, Appl
6	116	100.0	31	4	US-09-172-353-4	Sequence 4, Appli
7	116	100.0	32	3	US-09-105-678A-38	Sequence 38, Appl
8	116	100.0	33	3	US-09-105-678A-39	Sequence 39, Appl
9	111	95.7	20	3	US-09-105-678A-34	Sequence 34, Appl
10	111	95.7	21	3	US-09-105-678A-35	Sequence 35, Appl
11	111	95.7	22	3	US-09-105-678A-36	Sequence 36, Appl
12	111	95.7	31	3	US-09-105-678A-7	Sequence 7, Appli
13	111	95.7	31	3	US-09-105-678A-31	Sequence 31, Appl
14	111	95.7	32	3	US-09-105-678A-32	Sequence 32, Appl
15	111	95.7	33	3	US-09-105-678A-33	Sequence 33, Appl
16	105	90.5	19	3	US-09-105-678A-30	Sequence 30, Appl
17	105	90.5	20	3	US-09-105-678A-46	Sequence 46, Appl
18	105	90.5	21	3	US-09-105-678A-47	Sequence 47, Appl
19	105	90.5	22	3	US-09-105-678A-48	Sequence 48, Appl
20	105	90.5	31	3	US-09-105-678A-9	Sequence 9, Appli
21	105	90.5	31	3	US-09-105-678A-43	Sequence 43, Appl
22	105	90.5	32	3	US-09-105-678A-44	Sequence 44, Appl
23	105	90.5	33	3	US-09-105-678A-45	Sequence 45, Appl
24	104	89.7	21	3	US-09-105-678A-28	Sequence 28, Appl
25	100	86.2	29	3	US-09-105-678A-29	Sequence 29, Appl
26	50	43.1	349	1	US-08-118-270-71	Sequence 71, Appl
27	50	43.1	349	5	PCT-US93-08528-71	Sequence 71, Appl
28	49	42.2	774	3	US-08-902-632-2	Sequence 2, Appli
29	49	42.2	774	3	US-09-073-354-1	Sequence 1, Appli
30	49	42.2	774	3	US-08-656-005A-1	Sequence 1, Appli
31	49	42.2	774	4	US-09-073-259-1	Sequence 1, Appli
32	49	42.2	774	4	US-09-363-095-1	Sequence 1, Appli
33	46	39.7	330	2	US-08-815-176-1	Sequence 1, Appli
34	46	39.7	555	2	US-08-982-232-7	Sequence 7, Appli
35	46	39.7	555	2	US-08-982-232-14	Sequence 14, Appl
36	45	38.8	778	2	US-08-906-925-4	Sequence 4, Appli
37	45	38.8	779	1	US-08-375-134-12	Sequence 12, Appl
38	45	38.8	779	5	PCT-US95-15263-12	Sequence 12, Appl
39	45	38.8	940	4	US-09-078-347A-1	Sequence 1, Appli
40	44	37.9	1019	1	US-08-271-364A-7	Sequence 7, Appli
41	44	37.9	1019	2	US-08-222-715B-26	Sequence 26, Appl
42	43	37.1	274	4	US-09-141-821-3	Sequence 3, Appli
43	41	35.3	70	4	US-08-513-974B-34	Sequence 34, Appl
44	41	35.3	70	4	US-08-513-974B-317	Sequence 317, App
45	41	35.3	141	4	US-08-513-974B-320	Sequence 320, App

ALIGNMENTS

```
RESULT   1
US-09-105-678A-40
; Sequence 40, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-40


  Query Match           100.0%;  Score 116;  DB 3;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 4.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 TPDINPAWYTGRGIRPVGRF 20


RESULT   2
US-09-105-678A-41
; Sequence 41, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-41


  Query Match             100.0%;  Score 116;  DB 3;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 TPDINPAWYTGRGIRPVGRF 20


RESULT   3
US-09-105-678A-42
; Sequence 42, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-42


  Query Match             100.0%;  Score 116;  DB 3;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 TPDINPAWYTGRGIRPVGRF 20


RESULT   4
US-09-105-678A-8
; Sequence 8, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
```

```
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-8


  Query Match             100.0%;  Score 116;  DB 3;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYTGRGIRPVGRF 31


RESULT   5
US-09-105-678A-37
; Sequence 37, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-37


  Query Match             100.0%;  Score 116;  DB 3;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYTGRGIRPVGRF 31


RESULT   6
US-09-172-353-4
; Sequence 4, Application US/09172353
; Patent No. 6197530
; GENERAL INFORMATION:
;  APPLICANT: Stricker-Kongra, Alain
;  APPLICANT: Gu, Wei
;  TITLE OF INVENTION: GPR10 AS A TARGET FOR IDENTIFYING WEIGHT MODULATING COMPOUNDS
;  FILE REFERENCE: 07334/102001
;  CURRENT APPLICATION NUMBER: US/09/172,353
;  CURRENT FILING DATE: 1998-10-14
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 31
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-172-353-4


  Query Match             100.0%;  Score 116;  DB 4;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYTGRGIRPVGRF 31


RESULT   7
US-09-105-678A-38
; Sequence 38, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-38


  Query Match             100.0%;  Score 116;  DB 3;  Length 32;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYTGRGIRPVGRF 31


RESULT   8
US-09-105-678A-39
; Sequence 39, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-39


  Query Match             100.0%;  Score 116;  DB 3;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYTGRGIRPVGRF 31


RESULT   9
```

```
US-09-105-678A-34
; Sequence 34, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-34


  Query Match             95.7%;  Score 111;  DB 3;  Length 20;
  Best Local Similarity   95.0%;  Pred. No. 2.9e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          ||||||||| ||||||||||
Db      1 TPDINPAWYAGRGIRPVGRF 20


RESULT  10
US-09-105-678A-35
; Sequence 35, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-35


  Query Match             95.7%;  Score 111;  DB 3;  Length 21;
  Best Local Similarity   95.0%;  Pred. No. 3.1e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           |||||||||  ||||||||||
Db       1 TPDINPAWYAGRGIRPVGRF 20


RESULT  11
US-09-105-678A-36
; Sequence 36, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-36


  Query Match             95.7%;  Score 111;  DB 3;  Length 22;
  Best Local Similarity   95.0%;  Pred. No. 3.2e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           |||||||||  ||||||||||
Db       1 TPDINPAWYAGRGIRPVGRF 20


RESULT  12
US-09-105-678A-7
; Sequence 7, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-7


  Query Match             95.7%;  Score 111;  DB 3;  Length 31;
  Best Local Similarity   95.0%;  Pred. No. 4.7e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           |||||||||  ||||||||||
Db      12 TPDINPAWYAGRGIRPVGRF 31
```

```
RESULT  13
US-09-105-678A-31
; Sequence 31, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-31


  Query Match             95.7%;  Score 111;  DB 3;  Length 31;
  Best Local Similarity   95.0%;  Pred. No. 4.7e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           |||||||| ||||||||||
Db      12 TPDINPAWYAGRGIRPVGRF 31


RESULT  14
US-09-105-678A-32
; Sequence 32, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-32


  Query Match             95.7%;  Score 111;  DB 3;  Length 32;
  Best Local Similarity   95.0%;  Pred. No. 4.9e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           |||||||| ||||||||||
Db      12 TPDINPAWYAGRGIRPVGRF 31


RESULT  15
US-09-105-678A-33
; Sequence 33, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
```

```
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-33


  Query Match             95.7%;  Score 111;  DB 3;  Length 33;
  Best Local Similarity   95.0%;  Pred. No. 5.1e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           ||||||||| ||||||||||
Db      12 TPDINPAWYAGRGIRPVGRF 31


Search completed: April 17, 2001, 15:39:48
Job time: 317 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:48:48 ; Search time 39.1 Seconds
(without alignments)
17.522 Million cell updates/sec

Title: US-09-446-543A-50
Perfect score: 116
Sequence: 1 TPDINPAWYTGRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	116	100.0	83	1	PRRP_RAT	P81278 rattus norv
2	111	95.7	98	1	PRRP_BOVIN	P81264 bos taurus
3	105	90.5	87	1	PRRP_HUMAN	P81277 homo sapien
4	51	44.0	962	1	UVRA_METTH	O26543 methanobact
5	47	40.5	453	1	TBB2_GEOCN	P32925 geotrichum
6	47	40.5	719	1	ARP_YEAST	P32770 saccharomyc
7	47	40.5	972	1	UVRA_MYCTU	P94972 mycobacteri
8	47	40.5	973	1	UVRA_RHIME	P56899 rhizobium m
9	46.5	40.1	652	1	TETP_CLOPE	Q46306 clostridium
10	46	39.7	322	1	GRP2_MOUSE	O89100 m grb2-rela
11	46	39.7	330	1	GRP2_HUMAN	O75791 h grb2-rela
12	46	39.7	943	1	UVRA_HAEIN	P44410 haemophilus
13	45	38.8	482	1	R167_YEAST	P39743 saccharomyc
14	45	38.8	676	1	EXL1_HUMAN	Q92935 homo sapien
15	45	38.8	773	1	DPOL_THEGO	P56689 thermococcu
16	45	38.8	798	1	UNR_RAT	P18395 rattus norv
17	45	38.8	940	1	UVRA_ECOLI	P07671 escherichia
18	45	38.8	941	1	UVRA_SALTY	P37434 salmonella
19	45	38.8	953	1	UVRA_RICPR	Q9zcc3 rickettsia
20	45	38.8	1523	1	DPOL_THEFM	P74918 thermococcu
21	44	37.9	179	1	RK6_GUITH	O46908 guillardia
22	44	37.9	184	1	RL6_THEMA	Q9zae4 thermotoga
23	44	37.9	771	1	DPOL_PYRAB	P77916 pyrococcus
24	44	37.9	771	1	DPOL_PYRSE	P77932 pyrococcus
25	44	37.9	952	1	UVRA_THETH	Q56242 thermus aqu
26	44	37.9	970	1	UVRA_SYNY3	P73412 synechocyst
27	44	37.9	1014	1	UVRA_STRCO	Q9z507 streptomyce
28	44	37.9	1312	1	DPOL_PYRSD	Q51334 pyrococcus
29	43.5	37.5	1680	1	FUR2_DROME	P30432 drosophila
30	43	37.1	241	1	YG3D_YEAST	P53281 saccharomyc
31	43	37.1	272	1	LCB3_ROBPS	Q41160 robinia pse
32	43	37.1	476	1	YAAJ_ECOLI	P30143 escherichia
33	43	37.1	569	1	UVRA_VITST	Q08518 vitreoscill
34	43	37.1	950	1	UVRA_NEIGO	Q50968 neisseria g
35	43	37.1	960	1	UVRA_TREPA	O83527 treponema p
36	43	37.1	1282	1	TP2M_DICDI	P90520 dictyosteli
37	42.5	36.6	486	1	DMPC_PSESP	P19059 pseudomonas
38	42	36.2	247	1	MER5_ARATH	P24806 arabidopsis
39	42	36.2	272	1	TRA2_DROVI	O02008 drosophila
40	42	36.2	424	1	SAMB_SALTY	P23832 salmonella
41	42	36.2	926	1	UVRA_AQUAE	O66911 aquifex aeo
42	42	36.2	1235	1	DPOL_PYRHO	O59610 pyrococcus
43	41	35.3	101	1	VG30_BPML5	Q05239 mycobacteri
44	41	35.3	264	1	Y355_BUCAI	P57436 buchnera ap
45	41	35.3	342	1	Y762_METJA	Q58172 methanococc

ALIGNMENTS

RESULT 1
PRRP_RAT
ID PRRP_RAT STANDARD; PRT; 83 AA.
AC P81278;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE RELEASING PEPTIDE PRRP20].
GN PRH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RL Nature 393:272-276(1998).
CC -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC ----------

CC ----------
DR EMBL; AB015418; BAA29026.1; -.
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 21 BY SIMILARITY.
FT PEPTIDE 22 52 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT PEPTIDE 33 52 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 52 52 AMIDATION (G-53 PROVIDE AMIDE GROUP).
SQ SEQUENCE 83 AA; 9215 MW; D0C75A264EEE4F29 CRC64;

Query Match 100.0%; Score 116; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.8e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TPDINPAWYTGRGIRPVGRF 20
           ||||||||||||||||||||
Db      33 TPDINPAWYTGRGIRPVGRF 52
```

RESULT 2

```
PRRP_BOVIN
ID   PRRP_BOVIN     STANDARD;      PRT;    98 AA.
AC   P81264;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015417; BAA29025.1; -.
KW   Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     22
FT   PEPTIDE      23     53       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      33     53       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      53     53       AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   98 AA;  10544 MW;  08AC35A13B0FA908 CRC64;


  Query Match             95.7%;  Score 111;  DB 1;  Length 98;
  Best Local Similarity   95.0%;  Pred. No. 3.8e-10;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          |||||||| ||||||||||
Db     34 TPDINPAWYAGRGIRPVGRF 53


RESULT   3
PRRP_HUMAN
ID   PRRP_HUMAN     STANDARD;      PRT;    87 AA.
AC   P81277;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015419; BAA29027.1; -.
DR   MIM; 602663; -.
KW   Hormone; Amidation; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   PEPTIDE      23     53       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      34     53       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      53     53       AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   87 AA;  9639 MW;  229A2F3F50CF981B CRC64;


  Query Match             90.5%;  Score 105;  DB 1;  Length 87;
  Best Local Similarity   90.0%;  Pred. No. 2.8e-09;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYTGRGIRPVGRF 20
          |||||||| ||||||||||
Db     34 TPDINPAWYASRGIRPVGRF 53


RESULT   4
UVRA_METTH
ID   UVRA_METTH     STANDARD;      PRT;   962 AA.
AC   O26543;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR MTH443.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
```

```
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000828; AAB84949.1; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      38     45       ATP (POTENTIAL).
FT   NP_BIND     649    656       ATP (POTENTIAL).
FT   ZN_FING     748    774       C4-TYPE.
SQ   SEQUENCE   962 AA;  108395 MW;  2C0EF7FC41CCD060 CRC64;


  Query Match             44.0%;  Score 51;  DB 1;  Length 962;
  Best Local Similarity   40.5%;  Pred. No. 4.1;
  Matches   15;  Conservative    0;  Mismatches    4;  Indels   18;  Gaps    2;

Qy       1 TPDINPAWYTG----------------RGIRPVGRF 20
           ||  ||| |||                || || |||
Db     704 TPRSNPATYTGVFTHIRELFAQTPEARKRGYRP-GRF 739


RESULT   5
TBB2_GEOCN
ID   TBB2_GEOCN     STANDARD;      PRT;   453 AA.
AC   P32925;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   TUBULIN BETA-2 CHAIN.
OS   Geotrichum candidum (Oospora lactis).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Galactomyces.
OX   NCBI_TaxID=27317;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92079883; PubMed=1836049;
RA   Gold S.E., Casale W.L., Keen N.T.;
RT   "Characterization of two beta-tubulin genes from Geotrichum
RT   candidum.";
RL   Mol. Gen. Genet. 230:104-112(1991).
CC   -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
CC       BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
CC       CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC   -!- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
CC   -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S69627; AAB20557.1; ALT_SEQ.
DR   PIR; S18597; S18597.
DR   InterPro; IPR000217; -.
DR   InterPro; IPR002453; -.
DR   InterPro; IPR003008; -.
DR   Pfam; PF00091; tubulin; 1.
DR   PRINTS; PR01161; TUBULIN.
DR   PRINTS; PR01163; BETATUBULIN.
DR   PROSITE; PS00227; TUBULIN; 1.
DR   PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW   Microtubules; GTP-binding; Multigene family.
FT   NP_BIND     142    148       GTP (POTENTIAL).
SQ   SEQUENCE   453 AA;  50399 MW;  2EA9D8A0246E0371 CRC64;


  Query Match             40.5%;  Score 47;  DB 1;  Length 453;
  Best Local Similarity   70.0%;  Pred. No. 7.8;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       8 WYTGRGIRPV 17
           |||| |: ||
Db     399 WYTGEGMEPV 408


RESULT   6
ARP_YEAST
ID   ARP_YEAST      STANDARD;      PRT;   719 AA.
AC   P32770; Q12228;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ARP PROTEIN.
GN   ARP1 OR ARP OR YDL167C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AH22;
RX   MEDLINE=93247548; PubMed=8483449;
RA   Wehner E.P., Rao E., Brendel M.;
RT   "Molecular structure and genetic regulation of SFA, a gene
RT   responsible for resistance to formaldehyde in Saccharomyces
RT   cerevisiae, and characterization of its protein product.";
RL   Mol. Gen. Genet. 237:351-358(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RA   Pohl T.M.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 2 RANBP-TYPE ZINC FINGERS.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68020; CAA48159.1; -.
DR   EMBL; Z67750; CAA91579.1; -.
DR   EMBL; Z74215; CAA98741.1; -.
DR   PIR; S31139; S31139.
DR   HSSP; P04170; 6RXN.
DR   SGD; S0002326; NRP1.
DR   InterPro; IPR000504; -.
DR   InterPro; IPR001876; -.
DR   Pfam; PF00076; rrm; 1.
DR   Pfam; PF00641; zf-RanBP; 2.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW   Nuclear protein; Zinc-finger; RNA-binding; Repeat.
FT   DOMAIN      226    322       RNA-BINDING (RRM).
FT   DOMAIN      490    564       ASN-RICH.
FT   CONFLICT    493    493       I -> N (IN REF. 1).
SQ   SEQUENCE   719 AA;  79299 MW;  ADA9BC09FD582669 CRC64;


  Query Match             40.5%;  Score 47;  DB 1;  Length 719;
```

```
  Best Local Similarity   58.3%;  Pred. No. 12;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 AWYTGRGIRPVG 18
           :|:|  |:||||
Db     244 SWFTQYGVRPVG 255


RESULT   7
UVRA_MYCTU
ID   UVRA_MYCTU     STANDARD;      PRT;   972 AA.
AC   P94972;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR RV1638 OR MTCY06H11.02.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z85982; CAB06633.1; -.
DR   TubercuList; Rv1638; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      32     39       ATP (POTENTIAL).
FT   NP_BIND     654    661       ATP (POTENTIAL).
FT   ZN_FING     257    285       C4-TYPE (ATYPICAL).
FT   ZN_FING     753    779       C4-TYPE.
SQ   SEQUENCE   972 AA;  106131 MW;  8937A764E592D981 CRC64;


  Query Match             40.5%;  Score 47;  DB 1;  Length 972;
  Best Local Similarity   37.8%;  Pred. No. 16;
  Matches   14;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy       1 TPDINPAWYTG----------------RGIRPVGRF 20
             ||  ||| |||                || :| |||
Db     709 TPRSNPATYTGVFDKIRTLFAATTEAKVRGYQP-GRF 744


RESULT   8
UVRA_RHIME
ID   UVRA_RHIME     STANDARD;      PRT;   973 AA.
AC   P56899;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RA   Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
RA   Boistard P., Gouzy J., Kahn D., Thebault P., Goffeau A.,
RA   Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
RA   Vandenbol M., Puehler A., Becker A., Weidner S.;
RL   Submitted (MAY-2000) to the SWISS-PROT data bank.
RN   [2]
RP   SEQUENCE OF 1-140 FROM N.A.
RC   STRAIN=2021;
RX   MEDLINE=99430868; PubMed=10503543;
RA   Tapias A., Barbe J.;
RT   "Regulation of divergent transcription from the uvrA-ssb promoters in
RT   Sinorhizobium meliloti.";
RL   Mol. Gen. Genet. 262:121-130(1999).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF125162; AAF03210.1; -.
DR   InterPro; IPR001617; -.
DR   PROSITE; PS00211; ABC_TRANSPORTER; PARTIAL.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      34     41       ATP (POTENTIAL).
FT   NP_BIND     662    669       ATP (POTENTIAL).
FT   ZN_FING     761    787       C4-TYPE.
FT   CONFLICT     19     19       G -> A (IN REF. 2).
FT   CONFLICT     67     67       F -> S (IN REF. 2).
SQ   SEQUENCE   973 AA;  107191 MW;  3E1A8B14527A47FE CRC64;


  Query Match             40.5%;  Score 47;  DB 1;  Length 973;
  Best Local Similarity   37.8%;  Pred. No. 16;
  Matches   14;  Conservative    1;  Mismatches    4;  Indels   18;  Gaps    2;

Qy       1 TPDINPAWYTG----------------RGIRPVGRF 20
             ||  ||| |||                || :| |||
Db     717 TPRSNPATYTGAFTPIRDWFAGLPEAKARGYQP-GRF 752
```

```
RESULT   9
TETP_CLOPE
ID   TETP_CLOPE     STANDARD;      PRT;   652 AA.
AC   Q46306;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   TETRACYCLINE RESISTANCE PROTEIN TETP (TETB(P)).
GN   TETP OR TETB(P).
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CW92;
RX   MEDLINE=94224158; PubMed=8170402;
RA   Sloan J., McMurry L.M., Lyras D., Levy S.B., Rood J.I.;
RT   "The Clostridium perfringens Tet P determinant comprises two
RT   overlapping genes: tetA(P), which mediates active tetracycline
RT   efflux, and tetB(P), which is related to the ribosomal protection
RT   family of tetracycline-resistance determinants.";
RL   Mol. Microbiol. 11:403-415(1994).
CC   -!- FUNCTION: ABOLISH THE INHIBITORY EFFECT OF TETRACYCLIN ON PROTEIN
CC       SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       TETM/TETO SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L20800; AAA20117.1; -.
DR   InterPro; IPR000640; -.
DR   InterPro; IPR000795; -.
DR   InterPro; IPR002127; -.
DR   Pfam; PF00679; EFG_C; 1.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   PRINTS; PR01037; TCRTETOQM.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Protein biosynthesis; Antibiotic resistance; GTP-binding.
FT   NP_BIND      11     18       GTP (BY SIMILARITY).
FT   NP_BIND      75     79       GTP (BY SIMILARITY).
FT   NP_BIND     129    132       GTP (BY SIMILARITY).
SQ   SEQUENCE   652 AA;  72722 MW;  EFE5E1113D562113 CRC64;


  Query Match             40.1%;  Score 46.5;  DB 1;  Length 652;
  Best Local Similarity   55.6%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    6;  Indels    1;  Gaps    1;

Qy       3 DINPAWYT-GRGIRPVGR 19
           |:|| | | |  | | ||
Db     445 DLNPFWATVGLEIEPAGR 462


RESULT  10
GRP2_MOUSE
ID   GRP2_MOUSE     STANDARD;      PRT;   322 AA.
AC   O89100;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   GRB2-RELATED ADAPTOR PROTEIN 2 (GADS PROTEIN) (GROWTH FACTOR RECEPTOR
DE   BINDING PROTEIN) (GRBLG) (GRB-2-LIKE PROTEIN) (GRB2L) (HEMATOPOIETIC
DE   CELL-ASSOCIATED ADAPTOR PROTEIN GRPL) (GRB-2-RELATED MONOCYTIC ADAPTER
DE   PROTEIN) (MONOCYTIC ADAPTER) (MONA) (ADAPTER PROTEIN GRID).
GN   GRAP2 OR MONA OR GADS OR GRB2L OR GRID.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kedra D., Dumanski J.P.;
RT   "Cloning of the human and mouse growth factor receptor binding protein
RT   like genes.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bourette R.P., Arnaud S., Myles G.M., Rohrschneider L.R.,
RA   Blanchet J.P., Mouchiroud G.;
RT   "Mona, a novel hematopoietic-specific adaptor interacting with the
RT   macrophage-colony-stimulating factor receptor, is implicated in
RT   monocyte/macrophage development.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99087328; PubMed=9872323;
RA   Liu S.K., McGlade C.J.;
RT   "Gads is a novel SH2 and SH3 domain-containing adaptor protein that
RT   binds to tyrosine-phosphorylated Shc.";
RL   Oncogene 17:3073-3082(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Law C.-L., Ewings M.K., Chaudhary P.M., Solow S.A., Yun T.J.,
RA   Marshall A.J., Hood L., Clark E.A.;
RT   "GrpL, a Grb2-related adaptor protein, interacts with SLP-76 to
RT   regulate NF-AT activation.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Ellis J.H., Ashman C., Burden M., Kilpatrick K.E., Morse M.A.,
RA   Hamblin P.A.;
RT   "GRID, a novel Grb2-related adapter protein which interacts with the
RT   activated T cell co-stimulatory receptor CD28.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INTERACTS WITH SLP-76 TO REGULATE NF-AT ACTIVATION.
CC       BINDS TO TYROSINE-PHOSPHORYLATED SHC.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GRB2 / SEM-5 / DRK FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ011735; CAA09756.1; -.
DR   EMBL; AF055465; AAD08803.1; -.
DR   EMBL; AF053405; AAC98669.1; -.
DR   EMBL; AF129477; AAD41783.1; -.
DR   EMBL; AF236118; AAF60318.1; -.
DR   MGD; MGI:1333842; Mona.
DR   HSSP; P29354; 1GRI.
DR   InterPro; IPR000980; -.
DR   InterPro; IPR001452; -.
DR   Pfam; PF00017; SH2; 1.
DR   Pfam; PF00018; SH3; 2.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 2.
KW   SH2 domain; SH3 domain.
FT   DOMAIN       1     56       SH3.
FT   DOMAIN      58    149       SH2.
FT   DOMAIN     263    322       SH3.
SQ   SEQUENCE   322 AA;  36810 MW;  736311D0640CD3D0 CRC64;
```

```
  Query Match             39.7%;  Score 46;  DB 1;  Length 322;
  Best Local Similarity   50.0%;  Pred. No. 8;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5 NPAWYTGRGIRPVGRF 20
           ||:|:|||    :| |
Db     297 NPSWWTGRLHNKLGLF 312


RESULT  11
GRP2_HUMAN
ID   GRP2_HUMAN     STANDARD;      PRT;   330 AA.
AC   O75791; O43726;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   GRB2-RELATED ADAPTOR PROTEIN 2 (GADS PROTEIN) (GROWTH FACTOR RECEPTOR
DE   BINDING PROTEIN) (GRBLG) (GRF40 ADAPTOR PROTEIN) (GRB-2-LIKE PROTEIN)
DE   (GRB2L) (GRBX) (P38) (HEMATOPOIETIC CELL-ASSOCIATED ADAPTOR PROTEIN
DE   GRPL) (ADAPTER PROTEIN GRID).
GN   GRAP2 OR GADS OR GRB2L OR GRID.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RA   Frearson J.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kedra D., Dumanski J.P.;
RT   "Cloning of the human and mouse growth factor receptor binding protein
RT   like genes.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Asada H., Ishi N., Sasaki Y., Endo K., Kasai H., Tanaka N.,
RA   Takeshita T., Tsuchiya S., Konno T., Sugamura K.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99097254; PubMed=9878555;
RA   Qiu M., Hua S., Agrawal M., Li G., Cai J., Chan E., Zhou H., Luo Y.,
RA   Liu M.;
RT   "Molecular cloning and expression of human grap-2, a novel leukocyte-
RT   specific SH2- and SH3-containing adaptor-like protein that binds to
RT   gab-1.";
RL   Biochem. Biophys. Res. Commun. 253:443-447(1998).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Teramoto T., Nagashima M., Terai S., Thorgeirsson S.S.;
RT   "GrbX, new recruited signaling gene having homology with Grb2.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Law C.-L., Ewings M.K., Chaudhary P.M., Solow S.A., Yun T.J.,
RA   Marshall A.J., Hood L., Clark E.A.;
RT   "GrpL, a Grb2-related adaptor protein, interacts with SLP-76 to
RT   regulate NF-AT activation.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Ellis J.H., Ashman C., Burden M., Kilpatrick K.E., Morse M.A.,
RA   Hamblin P.A.;
RT   "GRID, a novel Grb2-related adapter protein which interacts with the
RT   activated T cell co-stimulatory receptor.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RA   Burgess J., Whiteley M.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INTERACTS WITH SLP-76 TO REGULATE NF-AT ACTIVATION.
CC       BINDS TO TYROSINE-PHOSPHORYLATED SHC.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GRB2 / SEM-5 / DRK FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y18051; CAA77021.1; -.
DR   EMBL; AJ011736; CAA09757.1; -.
DR   EMBL; AF042380; AAC69273.1; -.
DR   EMBL; AF102694; AAD04926.1; -.
DR   EMBL; AF090456; AAD13027.1; -.
DR   EMBL; AF129476; AAD41782.1; -.
DR   EMBL; AF236120; AAF60320.1; -.
DR   EMBL; Z82206; CAB05103.1; ALT_SEQ.
DR   HSSP; P29354; 1GRI.
DR   MIM; 604518; -.
DR   InterPro; IPR000980; -.
DR   InterPro; IPR001452; -.
DR   Pfam; PF00017; SH2; 1.
DR   Pfam; PF00018; SH3; 2.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 2.
KW   SH2 domain; SH3 domain.
FT   DOMAIN        1     56       SH3.
FT   DOMAIN       58    149       SH2.
FT   DOMAIN      271    330       SH3.
SQ   SEQUENCE   330 AA;  37909 MW;  74F4C8D0EBB56D55 CRC64;


  Query Match             39.7%;  Score 46;  DB 1;  Length 330;
  Best Local Similarity   50.0%;  Pred. No. 8.2;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5 NPAWYTGRGIRPVGRF 20
           ||:|:|||    :| |
Db     305 NPSWWTGRLHNKLGLF 320


RESULT  12
UVRA_HAEIN
ID   UVRA_HAEIN     STANDARD;      PRT;   943 AA.
AC   P44410; Q48151;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT A.
GN   UVRA OR HI0249.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
```

```
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NTHI N187;
RX   MEDLINE=97080495; PubMed=8921840;
RA   de la Morena M.L., Hendrixson D.R., St Geme J.W. III;
RT   "Isolation and characterization of the Haemophilus influenzae uvrA
RT   gene.";
RL   Gene 177:23-28(1996).
RN   [3]
RP   SEQUENCE OF 1-71 FROM N.A.
RC   STRAIN=NTHI TN106;
RX   MEDLINE=94341556; PubMed=8063092;
RA   Jarosik G.P., Hansen E.J.;
RT   "Cloning and sequencing of the Haemophilus influenzae ssb gene
RT   encoding single-strand DNA-binding protein.";
RL   Gene 146:101-103(1994).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC       AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC       STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32711; AAC21915.1; -.
DR   EMBL; U33877; AAC44592.1; -.
DR   EMBL; U04997; AAA60462.1; -.
DR   TIGR; HI0249; -.
DR   InterPro; IPR001617; -.
DR   Pfam; PF00005; ABC_tran; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW   DNA-binding; Zinc-finger.
FT   NP_BIND      31     38       ATP (POTENTIAL).
FT   NP_BIND     640    647       ATP (POTENTIAL).
FT   ZN_FING     253    280       C4-TYPE.
FT   ZN_FING     740    766       C4-TYPE.
FT   CONFLICT    163    163       V -> L (IN REF. 2).
FT   CONFLICT    236    236       E -> D (IN REF. 2).
FT   CONFLICT    425    425       R -> K (IN REF. 2).
FT   CONFLICT    463    463       I -> M (IN REF. 2).
FT   CONFLICT    514    514       E -> Q (IN REF. 2).
FT   CONFLICT    661    661       A -> T (IN REF. 2).
FT   CONFLICT    928    928       T -> E (IN REF. 2).
FT   CONFLICT    935    943       FLKPILEKP -> PLP (IN REF. 2).
SQ   SEQUENCE   943 AA;  104366 MW;  4DBA0DCFA602D465 CRC64;


  Query Match             39.7%;  Score 46;  DB 1;  Length 943;
  Best Local Similarity   37.8%;  Pred. No. 23;
  Matches   14;  Conservative    0;  Mismatches    5;  Indels   18;  Gaps    2;

Qy       1 TPDINPAWYTG----------------RGIRPVGRF 20
             ||  ||| |||                ||  | |||
Db     696 TPRSNPATYTGLFTPIRELFAGVPEARARGYNP-GRF 731


RESULT  13
R167_YEAST
ID   R167_YEAST     STANDARD;      PRT;   482 AA.
AC   P39743;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   REDUCED VIABILITY UPON STARVATION PROTEIN 167.
GN   RVS167 OR YDR388W OR D9509.8.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X2180;
RX   MEDLINE=93330299; PubMed=8336735;
RA   Bauer F., Urdaci M., Aigle M., Crouzet M.;
RT   "Alteration of a yeast SH3 protein leads to conditional viability
RT   with defects in cytoskeletal and budding patterns.";
RL   Mol. Cell. Biol. 13:5070-5084(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E.,
RA   Berno A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA   Hunicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA   Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA   Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA   Winant A., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   ACTIN-BINDING.
RX   MEDLINE=95236199; PubMed=7719850;
RA   Amberg D.C., Basart E., Botstein D.;
RT   "Defining protein interactions with yeast actin in vivo.";
RL   Nat. Struct. Biol. 2:28-35(1995).
CC   -!- FUNCTION: COMPONENT OF A CYTOSKELETAL STRUCTURE THAT IS REQUIRED
CC       FOR THE FORMATION OF ENDOCYTIC VESICLES AT THE PLASMA MEMBRANE
CC       LEVEL. COULD BE IMPLICATED IN CYTOSKELETAL REORGANIZATION IN
CC       RESPONSE TO ENVIRONMENTAL STRESSES AND COULD ACT IN THE BUDDING
CC       SITE SELECTION MECHANISM. BINDS TO ACTIN.
CC   -!- SIMILARITY: TO YEAST RVS161.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92092; AAA35051.1; -.
DR   EMBL; U32274; AAB64830.1; -.
DR   PIR; S40887; S40887.
DR   HSSP; P04002; 1ATF.
DR   SGD; S0002796; RVS167.
DR   InterPro; IPR001452; -.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PROSITE; PS50002; SH3; 1.
KW   Cytoskeleton; SH3 domain; Transmembrane; Actin-binding.
FT   DOMAIN      292    427       ALA/GLY/PRO-RICH.
FT   TRANSMEM    344    367       POTENTIAL.
FT   DOMAIN      421    482       SH3.
SQ   SEQUENCE   482 AA;  52774 MW;  3F0AB53EBCC95A5B CRC64;


  Query Match             38.8%;  Score 45;  DB 1;  Length 482;
  Best Local Similarity   47.6%;  Pred. No. 17;
  Matches   10;  Conservative    3;  Mismatches    2;  Indels    6;  Gaps    2;

Qy       1 TPDINPAWYTGR-----GIRP 16
             |||:|  |:|||     |: |
Db     454 TPDVN-EWWTGRYNGQQGVFP 473
```

```
RESULT  14
EXL1_HUMAN
ID   EXL1_HUMAN     STANDARD;      PRT;   676 AA.
AC   Q92935;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   EXOSTOSIN-LIKE 1 (EXOSTOSIN-L) (MULTIPLE EXOSTOSIS-LIKE PROTEIN).
GN   EXTL1 OR EXTL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97189339; PubMed=9037597;
RA   Wise C.A., Clines G.A., Massa H., Trask B.J., Lovett M.;
RT   "Identification and localization of the gene for EXTL, a third member
RT   of the multiple exostoses gene family.";
RL   Genome Res. 7:10-16(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Xu L., Deng H.X., Xia J.H., Pan Q., Liu C.Y.;
RT   "Mutations of the EXT genes in hereditary multiple exostoses in
RT   Chinese.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Wuyts W., Spieker N., Van Roy N., De Paepe A., De Boulle K.,
RA   Willems P.J., Van Hul W., Versteeg R., Speleman F.;
RT   "Refined physical mapping and genomic structure of the EXTL1 gene.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: APPEARS TO BE A TUMOR SUPPRESSOR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC
CC       RETICULUM (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67191; AAC51141.1; -.
DR   EMBL; AF083633; AAD02840.1; -.
DR   EMBL; AF083623; AAD02840.1; JOINED.
DR   EMBL; AF083624; AAD02840.1; JOINED.
DR   EMBL; AF083625; AAD02840.1; JOINED.
DR   EMBL; AF083626; AAD02840.1; JOINED.
DR   EMBL; AF083627; AAD02840.1; JOINED.
DR   EMBL; AF083628; AAD02840.1; JOINED.
DR   EMBL; AF083629; AAD02840.1; JOINED.
DR   EMBL; AF083630; AAD02840.1; JOINED.
DR   EMBL; AF083631; AAD02840.1; JOINED.
DR   EMBL; AF083632; AAD02840.1; JOINED.
DR   EMBL; AF153980; AAF73172.1; -.
DR   EMBL; AF151391; AAF73172.1; JOINED.
DR   MIM; 601738; -.
KW   Anti-oncogene; Multigene family; Transmembrane; Signal-anchor.
FT   TRANSMEM     10     30       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   676 AA;  74673 MW;  B5E006A8762E5633 CRC64;


  Query Match           38.8%;  Score 45;  DB 1;  Length 676;
  Best Local Similarity  45.0%;  Pred. No. 23;
  Matches    9;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYTGRGIRPVGRF 20
           :|   | :|  :| || |||
Db     400 SPQDFPFYYLQQGSRPEGRF 419


RESULT  15
DPOL_THEGO
ID   DPOL_THEGO     STANDARD;      PRT;   773 AA.
AC   P56689;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   DNA POLYMERASE (EC 2.7.7.7) (TO POL).
GN   POL OR POLA.
OS   Thermococcus gorgonarius.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX   NCBI_TaxID=71997;
RN   [1]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE=99199230; PubMed=10097083;
RA   Hopfner K.-P., Eichinger A., Engh R.A., Laue F., Ankenbauer W.,
RA   Huber R., Angerer B.;
RT   "Crystal structure of a thermostable type B DNA polymerase from
RT   Thermococcus gorgonarius.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:3600-3605(1999).
CC   -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC       EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC       N PYROPHOSPHATE + DNA(N).
CC   -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR   PDB; 1TGO; 22-MAR-99.
DR   InterPro; IPR002064; -.
DR   Pfam; PF00136; DNA_pol_B; 1.
DR   PRINTS; PR00106; DNAPOLB.
DR   PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   DNA-binding; Hydrolase; Nuclease; Exonuclease; Multifunctional enzyme;
KW   3D-structure.
FT   DISULFID    428    442
FT   DISULFID    506    509
SQ   SEQUENCE   773 AA;  89812 MW;  F67AF04E875FBE44 CRC64;


  Query Match           38.8%;  Score 45;  DB 1;  Length 773;
  Best Local Similarity  44.4%;  Pred. No. 26;
  Matches    8;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 DINPAWYTGRGIRPVGRF 20
           :|  || || |:  | |:
Db     294 EIAQAWETGEGLERVARY 311


Search completed: April 17, 2001, 15:48:49
Job time: 537 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:48:10 ; Search time 115.07 Seconds
(without alignments)
20.372 Million cell updates/sec

Title: US-09-446-543A-64
Perfect score: 113
Sequence: 1 TPDINPAWYASRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	84	74.3	117	13	Q9W624	Q9w624 carassius a
2	54	47.8	790	10	Q9M371	Q9m371 arabidopsis
3	51	45.1	54	4	Q9UJF9	Q9ujf9 homo sapien
4	51	45.1	465	4	O60687	O60687 homo sapien
5	47.5	42.0	333	2	Q9RJI0	Q9rji0 streptomyce
6	47	41.6	284	1	O50128	O50128 pyrococcus
7	47	41.6	333	2	Q9PH76	Q9ph76 xylella fas
8	47	41.6	938	10	Q9ZVE3	Q9zve3 arabidopsis
9	46.5	41.2	779	3	Q9P5J9	Q9p5j9 neurospora
10	46	40.7	1296	5	Q22452	Q22452 caenorhabdi
11	45.5	40.3	1501	10	Q9SD86	Q9sd86 arabidopsis
12	45	39.8	153	9	Q9T133	Q9t133 bacteriopha
13	45	39.8	207	5	Q27455	Q27455 brachycentr
14	45	39.8	390	2	Q9PH18	Q9ph18 xylella fas
15	45	39.8	419	4	Q9Y276	Q9y276 homo sapien
16	45	39.8	430	1	O27142	O27142 methanobact
17	45	39.8	540	10	Q9LGZ0	Q9lgz0 oryza sativ
18	45	39.8	767	5	Q20170	Q20170 caenorhabdi
19	44.5	39.4	259	2	Q9L1R4	Q9l1r4 streptomyce
20	44	38.9	276	2	O53929	O53929 mycobacteri
21	44	38.9	284	1	Q9V0Q3	Q9v0q3 pyrococcus
22	44	38.9	302	2	P96274	P96274 mycobacteri
23	44	38.9	309	5	O17234	O17234 caenorhabdi
24	44	38.9	428	2	Q9PKE7	Q9pke7 chlamydia m
25	44	38.9	527	5	O76383	O76383 caenorhabdi
26	43.5	38.5	374	2	O53319	O53319 mycobacteri
27	43	38.1	402	2	Q9KIW3	Q9kiw3 thiobacillu
28	43	38.1	466	2	O54042	O54042 pseudomonas
29	43	38.1	743	10	Q9SKV6	Q9skv6 arabidopsis
30	43	38.1	1359	10	Q9SW46	Q9sw46 arabidopsis
31	43	38.1	1364	10	Q9SW45	Q9sw45 arabidopsis
32	42.5	37.6	345	2	Q9K8V2	Q9k8v2 bacillus ha
33	42.5	37.6	1540	10	Q9SD90	Q9sd90 arabidopsis
34	42.5	37.6	3944	5	Q18667	Q18667 caenorhabdi
35	42	37.2	72	8	Q9TD69	Q9td69 mytilus cal
36	42	37.2	128	2	P74747	P74747 synechocyst
37	42	37.2	132	2	Q9PGJ6	Q9pgj6 xylella fas
38	42	37.2	184	2	O86838	O86838 streptomyce
39	42	37.2	216	2	Q9PBU7	Q9pbu7 xylella fas
40	42	37.2	219	10	Q9SL72	Q9sl72 arabidopsis
41	42	37.2	221	2	Q9RTS3	Q9rts3 deinococcus
42	42	37.2	232	2	Q9RZ21	Q9rz21 deinococcus
43	42	37.2	343	3	O74569	O74569 coprinus ci
44	42	37.2	459	3	O94284	O94284 schizosacch
45	42	37.2	512	3	O94269	O94269 schizosacch

ALIGNMENTS

RESULT 1
Q9W624
ID Q9W624 PRELIMINARY; PRT; 117 AA.
AC Q9W624;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE C-RF AMIDE PRECURSOR.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Satake H., Minakata H., Fujimoto M.;
RT "Carassius RFamide (C-RF amide).";
RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB020024; BAA76662.1; -.
SQ SEQUENCE 117 AA; 12879 MW; D5DC4CB22038C2B0 CRC64;

Query Match 74.3%; Score 84; DB 13; Length 117;
Best Local Similarity 60.0%; Pred. No. 1.2e-05;
Matches 12; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 TPDINPAWYASRGIRPVGRF 20
           :|:|:| ||  ||:||:|||
Db      56 SPEIDPFWYVGRGVRPIGRF 75
```

RESULT 2
Q9M371
ID Q9M371 PRELIMINARY; PRT; 790 AA.
AC Q9M371;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 87.4 KDA PROTEIN.
GN F15G16.60.
OS Arabidopsis thaliana (Mouse-ear cress).

```
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   De Haan M., Maarse A.C., Grivell L.A., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL132959; CAB71097.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   790 AA;  87375 MW;  B222724B75690F30 CRC64;


  Query Match             47.8%;  Score 54;  DB 10;  Length 790;
  Best Local Similarity   52.6%;  Pred. No. 3.9;
  Matches   10;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYASRGIRPVGRF 20
           |  ||  | |||::| ||:
Db     366 PPHNPRTYGSRGLQPHGRW 384


RESULT   3
Q9UJF9
ID   Q9UJF9      PRELIMINARY;      PRT;    54 AA.
AC   Q9UJF9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DJ479J7.3 (SUSHI-REPEAT PROTEIN (SRPUL)) (FRAGMENT).
GN   DJ479J7.3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lawlor S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035608; CAB55682.1; -.
FT   NON_TER      54     54
SQ   SEQUENCE   54 AA;  6110 MW;  E2F3C39F7B961A9F CRC64;


  Query Match             45.1%;  Score 51;  DB 4;  Length 54;
  Best Local Similarity   50.0%;  Pred. No. 0.64;
  Matches    8;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRP 16
           || : | |||  |  |
Db      18 TPAVTPTWYAGSGYYP 33


RESULT   4
O60687
ID   O60687      PRELIMINARY;      PRT;   465 AA.
AC   O60687;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   SUSHI-REPEAT PROTEIN.
GN   SRPUL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kurosawa H., Inukai T., Inaba T., Goi K., Chang K.-S., Sinjyo T.,
RA   Rakestraw K.M., Naeve C.W., Look T.A.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF060567; AAC15765.1; -.
DR   INTERPRO; IPR000436; -.
DR   INTERPRO; IPR001128; -.
DR   PFAM; PF00084; sushi; 3.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
SQ   SEQUENCE   465 AA;  52971 MW;  4D752B187FF3EFB8 CRC64;


  Query Match             45.1%;  Score 51;  DB 4;  Length 465;
  Best Local Similarity   50.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRP 16
           || : | |||  |  |
Db      18 TPAVTPTWYAGSGYYP 33


RESULT   5
Q9RJI0
ID   Q9RJI0      PRELIMINARY;      PRT;   333 AA.
AC   Q9RJI0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 36.3 KDA PROTEIN.
GN   SCF73.06C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL121746; CAB57411.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   333 AA;  36312 MW;  884FF2861837FD76 CRC64;


  Query Match             42.0%;  Score 47.5;  DB 2;  Length 333;
  Best Local Similarity   35.5%;  Pred. No. 16;
  Matches   11;  Conservative    3;  Mismatches    6;  Indels   11;  Gaps    1;

Qy       1 TPDINPAW----------YASRGIRPVGRF 20
           ||   |||           |: |:|| ||:
Db      19 TPWEEPAWRTEAIGWIEARLAAHGLRPTGRW 49


RESULT   6
O50128
ID   O50128      PRELIMINARY;      PRT;   284 AA.
AC   O50128;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 32.3 KDA PROTEIN PH1420.
```

```
GN   PH1420.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000006; BAA30526.1; -.
DR   INTERPRO; IPR001066; -.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   284 AA;  32319 MW;  8E0E7BC3711D3815 CRC64;


  Query Match             41.6%;  Score 47;  DB 1;  Length 284;
  Best Local Similarity   50.0%;  Pred. No. 16;
  Matches    9;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYASRGIRPVGR 19
           | | | :|| ||:  :||
Db     217 PYIEPTFYALRGLELLGR 234


RESULT   7
Q9PH76
ID   Q9PH76      PRELIMINARY;      PRT;   333 AA.
AC   Q9PH76;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYDROXYBENZOATE OCTAPRENYLTRANSFERASE.
GN   XF0068.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-157(2000).
DR   EMBL; AE003860; AAF82881.1; -.
DR   INTERPRO; IPR000537; -.
DR   PFAM; PF01040; UbiA; 1.
DR   PROSITE; PS00943; UBIA; UNKNOWN_1.
SQ   SEQUENCE   333 AA;  37931 MW;  ECF3F4716C962B95 CRC64;


  Query Match             41.6%;  Score 47;  DB 2;  Length 333;
  Best Local Similarity   53.3%;  Pred. No. 19;
  Matches    8;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 INPAWYASRGIRPVG 18
           ::| |  :|| ||||
Db      54 LDPYWKLARGDRPVG 68


RESULT   8
Q9ZVE3
ID   Q9ZVE3      PRELIMINARY;      PRT;   938 AA.
AC   Q9ZVE3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   T9F8.5 PROTEIN.
GN   T9F8.5.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA   Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
RT   "Arabidopsis thaliana chromosome II BAC T9F8 genomic sequence.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005561; AAC98465.1; -.
SQ   SEQUENCE   938 AA;  105542 MW;  2A42C66E9C590B78 CRC64;


  Query Match             41.6%;  Score 47;  DB 10;  Length 938;
  Best Local Similarity   66.7%;  Pred. No. 56;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5 NPAWYASRGIRPVGR 19
           |||| || ||  :||
Db      60 NPAWSASFGIFILGR 74


RESULT   9
Q9P5J9
ID   Q9P5J9      PRELIMINARY;      PRT;   779 AA.
AC   Q9P5J9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   RELATED TO BCS1 PROTEIN PRECURSOR.
GN   B23L21.300.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356172; CAB91698.1; -.
SQ   SEQUENCE   779 AA;  85796 MW;  CA7891402DFFBE30 CRC64;


  Query Match             41.2%;  Score 46.5;  DB 3;  Length 779;
  Best Local Similarity   58.8%;  Pred. No. 55;
  Matches   10;  Conservative    2;  Mismatches    2;  Indels    3;  Gaps    1;

Qy       1 TPDINPA---WYASRGI 14
           |  :|||   |||:|||
Db     286 TDYLNPATRRWYANRGI 302


RESULT  10
Q22452
ID   Q22452      PRELIMINARY;      PRT;  1296 AA.
AC   Q22452;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   SIMILAR TO AGRIN AND FOLLISTATIN.
GN   T13C2.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Du Z.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U40030; AAA81133.1; -.
DR   HSSP; P37109; 1PCE.
DR   INTERPRO; IPR000561; -.
DR   INTERPRO; IPR001239; -.
DR   INTERPRO; IPR001791; -.
DR   INTERPRO; IPR002049; -.
DR   INTERPRO; IPR002350; -.
DR   PFAM; PF00050; kazal; 9.
DR   PFAM; PF00053; laminin_EGF; 2.
DR   PFAM; PF00054; laminin_G; 1.
DR   PRINTS; PR00290; KAZALINHBTR.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
SQ   SEQUENCE   1296 AA;  145178 MW;  05094BC185839690 CRC64;


  Query Match             40.7%;  Score 46;  DB 5;  Length 1296;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+02;
  Matches   10;  Conservative    2;  Mismatches    6;  Indels    2;  Gaps    1;

Qy       1 TPDINPAWYASRGIRPVGRF 20
           | |||  || ||  : : ||
Db     975 TTDINSDWYFSR--KDINRF 992


RESULT  11
Q9SD86
ID   Q9SD86      PRELIMINARY;      PRT;  1501 AA.
AC   Q9SD86;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 164.4 KDA PROTEIN.
GN   F13G24.180.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA   Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133421; CAB62610.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1501 AA;  164360 MW;  EAD2AE0C9E8EE5D2 CRC64;


  Query Match             40.3%;  Score 45.5;  DB 10;  Length 1501;
  Best Local Similarity   34.5%;  Pred. No. 1.6e+02;
  Matches   10;  Conservative    6;  Mismatches    4;  Indels    9;  Gaps    2;

Qy       1 TPDINPAWYASRG------IRPV---GRF 20
           :| : |:||:  |      ::||   |||
Db    1184 SPQMAPSWYSQYGTFKNGLVQPVNDTGRF 1212


RESULT  12
Q9T133
ID   Q9T133      PRELIMINARY;      PRT;   153 AA.
AC   Q9T133;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENDONUCLEASE.
GN   3.
OS   Bacteriophage phi-YeO3-12.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae;
OC   T7-like phages.
OX   NCBI_TaxID=110457;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pajunen M.I., Kiljunen S.J., Skurnik M.;
RT   "Complete genomic sequence of the lytic bacteriophage phi-YeO3-12 of
RT   Yersinia enterocolitica serotype O:3.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ251805; CAB63604.1; -.
KW   Endonuclease.
SQ   SEQUENCE   153 AA;  17640 MW;  211571BBDE6C641D CRC64;


  Query Match             39.8%;  Score 45;  DB 9;  Length 153;
  Best Local Similarity   66.7%;  Pred. No. 17;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       9 YASRGIRPVGRF 20
           ||:||:| || |
```

```
Db      5 YAARGVRKVGAF 16


RESULT  13
Q27455
ID   Q27455      PRELIMINARY;      PRT;   207 AA.
AC   Q27455;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PHOSPHOENOLPYRUVATE CARBOXYKINASE (EC 4.1.1.32)
DE   (PHOSPHOENOLPYRUVATE CARBOXYKINASE (GTP))
DE   (PHOSPHOENOLPYRUVATE CARBOXYLASE) (PHOSPHOPYRUVATE CARBOXYLASE)
DE   (PEPCK) (FRAGMENT).
OS   Brachycentrus nigrosoma.
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Trichoptera; Integripalpia;
OC   Limnephiloidea; Brachycentridae; Brachycentrus.
OX   NCBI_TaxID=41036;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97036855; PubMed=8882502;
RA   Friedlander T.P., Regier J.C., Mitter C., Wagner D.L.;
RT   "A nuclear gene for higher level phylogenetics: phosphoenolpyruvate
RT   carboxykinase tracks mesozoic-age divergences within Lepidoptera
RT   (Insecta).";
RL   Mol. Biol. Evol. 13:594-604(1996).
DR   EMBL; U28445; AAB40362.1; -.
DR   INTERPRO; IPR000364; -.
DR   PFAM; PF00821; PEPCK; 1.
DR   PROSITE; PS00505; PEPCK_GTP; 1.
KW   Pyruvate; Lyase.
FT   NON_TER       1      1
FT   NON_TER     207    207
SQ   SEQUENCE   207 AA;  22123 MW;  3488D5E47B56B3E8 CRC64;


  Query Match           39.8%;  Score 45;  DB 5;  Length 207;
  Best Local Similarity  37.0%;  Pred. No. 23;
  Matches   10;  Conservative    3;  Mismatches    4;  Indels   10;  Gaps    1;

Qy       2 PDINPAWYASRGI----------RPVG 18
           | |:||| :| |:          || |
Db     146 PTIDPAWESSAGVPISAILFGGRRPQG 172


RESULT  14
Q9PH18
ID   Q9PH18      PRELIMINARY;      PRT;   390 AA.
AC   Q9PH18;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CYSTEINE SYNTHASE.
GN   XF0128.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-157(2000).
DR   EMBL; AE003866; AAF82941.1; -.
DR   INTERPRO; IPR001926; -.
DR   PFAM; PF00291; PALP; 1.
SQ   SEQUENCE   390 AA;  42479 MW;  D294072E6A55188A CRC64;


  Query Match           39.8%;  Score 45;  DB 2;  Length 390;
  Best Local Similarity  70.0%;  Pred. No. 45;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 NPAWYASRGI 14
           :|||||: ||
Db     354 SPAWYAAHGI 363


RESULT  15
Q9Y276
ID   Q9Y276      PRELIMINARY;      PRT;   419 AA.
AC   Q9Y276;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   H-BCS1.
GN   BCS1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99097350; PubMed=9878253;
RA   Petruzzella V., Tiranti V., Fernandez P., Ianna P., Carrozzo R.,
RA   Zeviani M.;
RT   "Identification and characterization of human cDNAs specific to BCS1,
RT   PET112, SCO1, COX15, and COX11, five genes involved in the formation
RT   and function of the mitochondrial respiratory chain.";
RL   Genomics 54:494-504(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=96207227; PubMed=8619474;
RA   Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
RT   "A 'double adaptor' method for improved shotgun library
RT   construction.";
RL   Anal. Biochem. 236:107-113(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=97264341; PubMed=9110174;
RA   Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
RA   Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
RT   "Large-scale concatenation cDNA sequencing.";
RL   Genome Res. 7:353-358(1997).
DR   EMBL; AF026849; AAD08638.1; -.
```

DR EMBL; AF038195; AAB97365.1; -.
DR INTERPRO; IPR001939; -.
DR PFAM; PF00004; AAA; 1.
KW Hypothetical protein.
SQ SEQUENCE 419 AA; 47534 MW; 7F0F98BA62F2CBB8 CRC64;

Query Match 39.8%; Score 45; DB 4; Length 419;
Best Local Similarity 70.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       5 NPAWYASRGI 14
           || ||  |||
Db     211 NPKWYTDRGI 220
```

Search completed: April 17, 2001, 15:48:11
Job time: 564 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:45:57 ; Search time 70.08 Seconds
(without alignments)
19.613 Million cell updates/sec

Title: US-09-446-543A-64
Perfect score: 113
Sequence: 1 TPDINPAWYASRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	54	47.8	790	2	T47959	hypothetical prote
2	50.5	44.7	664	2	F83376	conserved hypothet
3	47	41.6	284	2	F71015	hypothetical prote
4	47	41.6	333	2	H82852	hydroxybenzoate oc
5	46.5	41.2	779	2	T49717	related to BCS1 pr
6	46	40.7	1296	2	T16859	hypothetical prote
7	45.5	40.3	1501	2	T45623	hypothetical prote
8	45	39.8	390	2	G82844	cysteine synthase
9	45	39.8	430	1	B69009	conserved hypothet
10	45	39.8	767	2	T21969	hypothetical prote
11	44	38.9	276	2	D70817	hypothetical prote
12	44	38.9	284	2	A75117	hypothetical prote
13	44	38.9	302	2	B70631	hypothetical prote
14	44	38.9	309	2	T32376	hypothetical prote
15	44	38.9	428	2	F81694	pyruvate dehydroge
16	44	38.9	476	2	G64720	probable amino aci
17	44	38.9	527	2	T33175	hypothetical prote
18	44	38.9	719	2	S61046	ARP1 protein - yea
19	43.5	38.5	374	2	G70947	hypothetical prote
20	43	38.1	220	2	C83292	probable glutathio
21	43	38.1	250	2	G83400	hypothetical prote
22	43	38.1	376	2	A48197	opsin, ocellar - A
23	43	38.1	376	2	B48197	opsin, lateral eye
24	43	38.1	468	2	C83160	nitrite extrusion
25	43	38.1	1359	2	T10235	xanthine dehydroge
26	43	38.1	1364	2	T10236	xanthine dehydroge
27	42.5	37.6	345	2	D84012	N-acetylglutamate
28	42.5	37.6	1540	2	T45619	hypothetical prote
29	42.5	37.6	3944	2	T19997	hypothetical prote
30	42	37.2	128	2	S76955	hypothetical prote
31	42	37.2	132	2	B82824	protein-export mem
32	42	37.2	184	2	T35841	probable membrane
33	42	37.2	216	2	H82608	transcription regu
34	42	37.2	221	2	A75366	probable transcrip
35	42	37.2	232	2	G75608	hypothetical prote
36	42	37.2	311	2	S66600	cytochrome-c oxida
37	42	37.2	364	2	B64766	yaiW protein - Esc
38	42	37.2	367	2	G83309	conserved hypothet
39	42	37.2	459	2	T43558	sulfide dehydrogen
40	42	37.2	512	2	T40815	probable ubiquitin
41	42	37.2	546	2	A32260	cholesterol oxidas
42	42	37.2	547	2	T48551	fructosidase-like
43	42	37.2	581	2	A42743	pol polyprotein -
44	42	37.2	677	2	T07944	probable long-chai
45	42	37.2	843	1	GNVWK	pol polyprotein -

ALIGNMENTS

RESULT 1
T47959
hypothetical protein F15G16.60 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47959
R;De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Q
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z24480
A;Accession: T47959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-790 <DEH>
A;Cross-references: EMBL:AL132959
A;Experimental source: cultivar Columbia; BAC clone F15G16
C;Genetics:
A;Map position: 3
A;Introns: 39/1; 678/2; 698/3; 773/2
A;Note: F15G16.60

Query Match 47.8%; Score 54; DB 2; Length 790;
Best Local Similarity 52.6%; Pred. No. 1.9;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```
Qy       2 PDINPAWYASRGIRPVGRF 20
           |  ||  | |||::| ||:
Db     366 PPHNPRTYGSRGLQPHGRW 384
```

RESULT 2
F83376
conserved hypothetical protein PA2151 [imported] - Pseudomonas aeruginosa (strain PAO
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83376
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: F83376
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-664 <STO>
A;Cross-references: GB:AE004642; GB:AE004091; NID:g9948163; PIDN:AAG05539.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2151

```
  Query Match             44.7%;  Score 50.5;  DB 2;  Length 664;
  Best Local Similarity   62.5%;  Pred. No. 5.5;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    1;  Gaps    1;

Qy      1 TPDINPAWYASRGIRP 16
          |||||| |:  |  ||
Db    478 TPDINP-WFLQRSGRP 492


RESULT   3
F71015
hypothetical protein PH1420 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: F71015
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: F71015
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-284 <KAW>
A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30526.1; PID:g3257843
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1420
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1420


  Query Match             41.6%;  Score 47;  DB 2;  Length 284;
  Best Local Similarity   50.0%;  Pred. No. 8;
  Matches    9;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      2 PDINPAWYASRGIRPVGR 19
          | | | :|| ||:  :||
Db    217 PYIEPTFYALRGLELLGR 234


RESULT   4
H82852
hydroxybenzoate octaprenyltransferase XF0068 [imported] - Xylella fastidiosa (strain 9a5
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
C;Accession: H82852
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H82852
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-333 <SIM>
A;Cross-references: GB:AE003860; GB:AE003849; NID:g9104830; PIDN:AAF82881.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Content: annotation
C;Genetics:
A;Gene: XF0068
C;Superfamily: 4-hydroxybenzoate octaprenyltransferase


  Query Match             41.6%;  Score 47;  DB 2;  Length 333;
  Best Local Similarity   53.3%;  Pred. No. 9.4;
  Matches    8;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      4 INPAWYASRGIRPVG 18
          ::| |  :|| ||||
Db     54 LDPYWKLARGDRPVG 68


RESULT   5
T49717
related to BCS1 protein precursor [imported] - Neurospora crassa
N;Alternate names: protein B23L21.300
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49717
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49717
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-779 <SCH>
A;Cross-references: EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.300
A;Experimental source: BAC clone B23L21; strain OR74A
C;Genetics:
A;Gene: NCSP:B23L21.300
A;Map position: 6


  Query Match             41.2%;  Score 46.5;  DB 2;  Length 779;
  Best Local Similarity   58.8%;  Pred. No. 27;
  Matches   10;  Conservative    2;  Mismatches    2;  Indels    3;  Gaps    1;

Qy      1 TPDINPA---WYASRGI 14
          |  :|||   |||:|||
Db    286 TDYLNPATRRWYANRGI 302


RESULT   6
T16859
hypothetical protein T13C2.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C;Accession: T16859
R;Du, Z.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid T13C2.
A;Reference number: Z18591
A;Accession: T16859
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <DUZ>
A;Cross-references: EMBL:U40030; NID:g1055164; PID:g1055165; PIDN:AAA81133.1; CESP:T1
C;Genetics:
A;Gene: CESP:T13C2.5
A;Introns: 62/3; 93/3; 124/3; 142/1; 212/1; 293/1; 336/2; 406/3; 437/1; 508/1; 584/1;


  Query Match             40.7%;  Score 46;  DB 2;  Length 1296;
  Best Local Similarity   50.0%;  Pred. No. 54;
  Matches   10;  Conservative    2;  Mismatches    6;  Indels    2;  Gaps    1;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          | |||  || ||  : : ||
Db    975 TTDINSDWYFSR--KDINRF 992


RESULT   7
```

T45623
hypothetical protein F13G24.180 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T45623
R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23009
A;Accession: T45623
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1501 <BEV>
A;Cross-references: EMBL:AL133421
A;Experimental source: cultivar Columbia; BAC clone F13G24
C;Genetics:
A;Map position: 5
A;Introns: 64/1; 739/3; 785/2; 1302/2; 1318/3; 1399/2; 1434/2
A;Note: F13G24.180

Query Match 40.3%; Score 45.5; DB 2; Length 1501;
Best Local Similarity 34.5%; Pred. No. 75;
Matches 10; Conservative 6; Mismatches 4; Indels 9; Gaps 2;

Qy 1 TPDINPAWYASRG------IRPV---GRF 20
:| : |:||: | ::|| |||
Db 1184 SPQMAPSWYSQYGTFKNGLVQPVNDTGRF 1212

RESULT 8
G82844
cysteine synthase XF0128 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: G82844
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: G82844
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-390 <SIM>
A;Cross-references: GB:AE003866; GB:AE003849; NID:g9104906; PIDN:AAF82941.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0128

Query Match 39.8%; Score 45; DB 2; Length 390;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 NPAWYASRGI 14
:|||||: ||
Db 354 SPAWYAAHGI 363

RESULT 9
B69009
conserved hypothetical protein MTH1070 - Methanobacterium thermoautotrophicum (strain
C;Species: Methanobacterium thermoautotrophicum
C;Date: 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change 23-Jul-1999
C;Accession: B69009
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: B69009
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-430 <MTH>
A;Cross-references: GB:AE000877; GB:AE000666; NID:g2622157; PIDN:AAB85559.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1070
C;Superfamily: conserved hypothetical protein MTH1070

Query Match 39.8%; Score 45; DB 1; Length 430;
Best Local Similarity 61.5%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 DINPAWYASRGIR 15
|||| | | | |
Db 191 DINPEWVAGRACR 203

RESULT 10
T21969
hypothetical protein F38E11.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21969
R;Matthews, P.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19495
A;Accession: T21969
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-767 <WIL>
A;Cross-references: EMBL:Z68342; PIDN:CAA92775.1; GSPDB:GN00022; CESP:F38E11.7
A;Experimental source: clone F38E11
C;Genetics:
A;Gene: CESP:F38E11.7
A;Map position: 4
A;Introns: 50/2; 118/1; 139/2; 189/3; 226/1; 248/1; 287/2; 375/2; 432/3; 465/3; 548/1

Query Match 39.8%; Score 45; DB 2; Length 767;
Best Local Similarity 75.0%; Pred. No. 45;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PDINPAWY 9
||: ||||
Db 747 PDVKPAWY 754

RESULT 11
D70817
hypothetical protein Rv1716 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70817
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

```
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: D70817
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-276 <COL>
A;Cross-references: GB:AL022003; GB:AL123456; NID:g3261547; PIDN:CAA17613.1; PID:g326154
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1716


  Query Match             38.9%;  Score 44;  DB 2;  Length 276;
  Best Local Similarity   20.0%;  Pred. No. 23;
  Matches   10;  Conservative    5;  Mismatches    3;  Indels   32;  Gaps    1;

Qy       1 TPDINPA-------------------------------WYASRGIRPVG 18
             |||| |                               |:|::|:: ||
Db     111 TPDIRPGDIVVVNTGWHHKYADSAEYYAYSPGFDKKAGEWFAAKGVKAVG 160


RESULT  12
A75117
hypothetical protein PAB0502 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: A75117
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: A75117
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-284 <KAW>
A;Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49650.1; PID:g545816
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB0502
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1420


  Query Match             38.9%;  Score 44;  DB 2;  Length 284;
  Best Local Similarity   44.4%;  Pred. No. 23;
  Matches    8;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYASRGIRPVGR 19
           | | | :|| ||:  :|:
Db     217 PYIEPTFYALRGLELLGK 234


RESULT  13
B70631
hypothetical protein Rv0428c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: B70631
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: B70631
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-302 <COL>
A;Cross-references: GB:Z84724; GB:AL123456; NID:g3261708; PIDN:CAB06568.1; PID:g1817694
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0428c
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0428c


  Query Match             38.9%;  Score 44;  DB 2;  Length 302;
  Best Local Similarity   70.0%;  Pred. No. 25;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       7 AWYASRGIRP 16
             |||||| ::|
Db     126 AWYASRDLQP 135


RESULT  14
T32376
hypothetical protein K10F12.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32376
R;Wohldmann, P.; Beck, C.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid K10F12.
A;Reference number: Z21157
A;Accession: T32376
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-309 <WOH>
A;Cross-references: EMBL:AF025462; PIDN:AAB71002.1; GSPDB:GN00021; CESP:K10F12.4
A;Experimental source: strain Bristol N2; clone K10F12
C;Genetics:
A;Gene: CESP:K10F12.4
A;Map position: 3
A;Introns: 31/3; 123/2; 196/3; 239/1


  Query Match             38.9%;  Score 44;  DB 2;  Length 309;
  Best Local Similarity   44.4%;  Pred. No. 25;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    2;  Gaps    1;

Qy       2 PDINPAWYASRGIRPVGR 19
           || :| ||  :   |:||
Db     131 PDRSPNWYLPKS--PIGR 146


RESULT  15
F81694
pyruvate dehydrogenase, E2 component, dihydrolipoamide acetyltransferase, probable TC
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C;Accession: F81694
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: A81500; MUID:20150255
A;Accession: F81694
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-428 <TET>
A;Cross-references: GB:AE002320; GB:AE002160; NID:g7190551; PIDN:AAF39360.1; PID:g719
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0518
C;Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology


  Query Match             38.9%;  Score 44;  DB 2;  Length 428;
  Best Local Similarity   37.5%;  Pred. No. 36;
  Matches    6;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYASRGIRPV 17
           ||:||  |    : |:
Db     196 PDVNPGSYVEESLSPI 211
```

Search completed: April 17, 2001, 15:45:58
Job time: 606 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:48:51 ; Search time 39.1 Seconds
(without alignments)
17.522 Million cell updates/sec

Title: US-09-446-543A-64
Perfect score: 113
Sequence: 1 TPDINPAWYASRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	113	100.0	87	1	PRRP_HUMAN	P81277 homo sapien
2	109	96.5	98	1	PRRP_BOVIN	P81264 bos taurus
3	105	92.9	83	1	PRRP_RAT	P81278 rattus norv
4	47	41.6	676	1	EXL1_HUMAN	Q92935 homo sapien
5	46	40.7	383	1	CYCR_CHRVI	O82947 chromatium
6	44	38.9	476	1	YAAJ_ECOLI	P30143 escherichia
7	44	38.9	719	1	ARP_YEAST	P32770 saccharomyc
8	43.5	38.5	860	1	VG12_BPB03	Q37893 bacteriopha
9	43	38.1	376	1	OPS1_LIMPO	P35360 limulus pol
10	43	38.1	376	1	OPS2_LIMPO	P35361 limulus pol
11	42	37.2	332	1	LYTB_MYCLE	Q9x781 mycobacteri
12	42	37.2	364	1	YAIW_ECOLI	P77562 escherichia
13	42	37.2	546	1	CHOD_STRSQ	P12676 streptomyce
14	42	37.2	581	1	POL_MLVRK	P31795 radiation m
15	42	37.2	843	1	POL_MLVAK	P03357 akr murine
16	42	37.2	1196	1	POL_MLVAV	P03356 akv murine
17	42	37.2	1196	1	POL_MLVRD	P11227 radiation m
18	41.5	36.7	265	1	UBIE_RICPR	Q9zcp3 rickettsia
19	41	36.3	149	1	ENRN_BPT7	P00641 bacteriopha
20	41	36.3	342	1	Y762_METJA	Q58172 methanococc
21	41	36.3	347	1	Y576_METJA	Q57996 methanococc
22	41	36.3	379	1	YXI1_ANASP	P29978 anabaena sp
23	41	36.3	622	1	PPCC_RAT	P07379 rattus norv
24	41	36.3	986	1	CYGR_ARBPU	P11528 arbacia pun
25	40.5	35.8	751	1	PA26_RAT	P97570 rattus norv
26	40	35.4	158	1	RL15_AERPE	Q9yf98 aeropyrum p
27	40	35.4	181	1	YK69_CAEEL	P34347 caenorhabdi
28	40	35.4	424	1	SAMB_SALTY	P23832 salmonella
29	40	35.4	556	1	DPOL_WHVW6	P11292 woodchuck h
30	40	35.4	580	1	ATPU_YEAST	P22136 saccharomyc
31	40	35.4	625	1	DHGL_DROME	P18173 drosophila
32	40	35.4	746	1	EXT1_HUMAN	Q16394 homo sapien
33	40	35.4	746	1	EXT1_MOUSE	P97464 mus musculu
34	40	35.4	788	1	REC2_HAEIN	P44408 haemophilus
35	40	35.4	792	1	SYFB_XYLFA	Q9pfd6 xylella fas
36	40	35.4	879	1	DPOL_WHV1	P03160 woodchuck h
37	40	35.4	884	1	DPOL_WHV59	P12899 woodchuck h
38	40	35.4	884	1	DPOL_WHV7	P12898 woodchuck h
39	40	35.4	962	1	UVRA_METTH	O26543 methanobact
40	40	35.4	973	1	UVRA_RHIME	P56899 rhizobium m
41	40	35.4	987	1	EPB4_MOUSE	P54761 mus musculu
42	39.5	35.0	831	1	DPO1_THEFL	P30313 thermus aqu
43	39.5	35.0	1122	1	TERT_MOUSE	O70372 mus musculu
44	39.5	35.0	1882	1	Y468_MYCPN	P75109 mycoplasma
45	39	34.5	61	1	YLCE_ECOLI	P77087 escherichia

ALIGNMENTS

```
RESULT   1
PRRP_HUMAN
ID   PRRP_HUMAN     STANDARD;      PRT;    87 AA.
AC   P81277;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015419; BAA29027.1; -.
DR   MIM; 602663; -.
KW   Hormone; Amidation; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   PEPTIDE      23     53       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      34     53       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      53     53       AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   87 AA;  9639 MW;  229A2F3F50CF981B CRC64;


  Query Match          100.0%;  Score 113;  DB 1;  Length 87;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db     34 TPDINPAWYASRGIRPVGRF 53
```

```
RESULT   2
PRRP_BOVIN
ID   PRRP_BOVIN     STANDARD;      PRT;    98 AA.
AC   P81264;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015417; BAA29025.1; -.
KW   Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     22
FT   PEPTIDE      23     53       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      33     53       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      53     53       AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   98 AA;  10544 MW;  08AC35A13B0FA908 CRC64;


  Query Match             96.5%;  Score 109;  DB 1;  Length 98;
  Best Local Similarity   95.0%;  Pred. No. 2.2e-10;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPVGRF 20
           |||||||||| |||||||||
Db      34 TPDINPAWYAGRGIRPVGRF 53


RESULT   3
PRRP_RAT
ID   PRRP_RAT       STANDARD;      PRT;    83 AA.
AC   P81278;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE   HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE   RELEASING PEPTIDE PRRP20].
GN   PRH.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98268781; PubMed=9607765;
RA   Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA   Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA   Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT   "A prolactin-releasing peptide in the brain.";
RL   Nature 393:272-276(1998).
CC   -!- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC       EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HGR3. MAY STIMULATE
CC       LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC   -!- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015418; BAA29026.1; -.
KW   Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     21       BY SIMILARITY.
FT   PEPTIDE      22     52       PROLACTIN-RELEASING PEPTIDE PRRP31.
FT   PEPTIDE      33     52       PROLACTIN-RELEASING PEPTIDE PRRP20.
FT   MOD_RES      52     52       AMIDATION (G-53 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   83 AA;  9215 MW;  D0C75A264EEE4F29 CRC64;


  Query Match             92.9%;  Score 105;  DB 1;  Length 83;
  Best Local Similarity   90.0%;  Pred. No. 7.8e-10;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPVGRF 20
           |||||||||  |||||||||
Db      33 TPDINPAWYTGRGIRPVGRF 52


RESULT   4
EXL1_HUMAN
ID   EXL1_HUMAN     STANDARD;      PRT;   676 AA.
AC   Q92935;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   EXOSTOSIN-LIKE 1 (EXOSTOSIN-L) (MULTIPLE EXOSTOSIS-LIKE PROTEIN).
GN   EXTL1 OR EXTL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97189339; PubMed=9037597;
RA   Wise C.A., Clines G.A., Massa H., Trask B.J., Lovett M.;
RT   "Identification and localization of the gene for EXTL, a third member
RT   of the multiple exostoses gene family.";
RL   Genome Res. 7:10-16(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Xu L., Deng H.X., Xia J.H., Pan Q., Liu C.Y.;
RT   "Mutations of the EXT genes in hereditary multiple exostoses in
RT   Chinese.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Wuyts W., Spieker N., Van Roy N., De Paepe A., De Boulle K.,
RA   Willems P.J., Van Hul W., Versteeg R., Speleman F.;
RT   "Refined physical mapping and genomic structure of the EXTL1 gene.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: APPEARS TO BE A TUMOR SUPPRESSOR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC
CC       RETICULUM (BY SIMILARITY).
```

```
CC   -!- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67191; AAC51141.1; -.
DR   EMBL; AF083633; AAD02840.1; -.
DR   EMBL; AF083623; AAD02840.1; JOINED.
DR   EMBL; AF083624; AAD02840.1; JOINED.
DR   EMBL; AF083625; AAD02840.1; JOINED.
DR   EMBL; AF083626; AAD02840.1; JOINED.
DR   EMBL; AF083627; AAD02840.1; JOINED.
DR   EMBL; AF083628; AAD02840.1; JOINED.
DR   EMBL; AF083629; AAD02840.1; JOINED.
DR   EMBL; AF083630; AAD02840.1; JOINED.
DR   EMBL; AF083631; AAD02840.1; JOINED.
DR   EMBL; AF083632; AAD02840.1; JOINED.
DR   EMBL; AF153980; AAF73172.1; -.
DR   EMBL; AF151391; AAF73172.1; JOINED.
DR   MIM; 601738; -.
KW   Anti-oncogene; Multigene family; Transmembrane; Signal-anchor.
FT   TRANSMEM     10     30       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   676 AA;  74673 MW;  B5E006A8762E5633 CRC64;


  Query Match             41.6%;  Score 47;  DB 1;  Length 676;
  Best Local Similarity   45.0%;  Pred. No. 7.8;
  Matches    9;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPVGRF 20
           :|   | :|  :| || |||
Db     400 SPQDFPFYYLQQGSRPEGRF 419


RESULT   5
CYCR_CHRVI
ID   CYCR_CHRVI     STANDARD;      PRT;   383 AA.
AC   O82947;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PHOTOSYNTHETIC REACTION CENTER CYTOCHROME C SUBUNIT PRECURSOR.
GN   PUFC.
OS   Chromatium vinosum.
OC   Bacteria; Proteobacteria; gamma subdivision; Chromatiaceae;
OC   Allochromatium.
OX   NCBI_TaxID=1049;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D;
RA   Corson G.E., Nagashima K.V., Matsuura K., Sakuragi Y., Ruwanthi W.,
RA   Qin H., Allen R., Knaff D.B.;
RT   "Primary structure of genes encoding light-harvesting and reaction
RT   center proteins from Chromatium vinosum.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE REACTION CENTER OF PURPLE BACTERIA CONTAIN A
CC       TIGHTLY BOUND CYTOCHROME MOLECULE WHICH RE-REDUCES THE PHOTO
CC       OXIDIZED PRIMARY ELECTRON DONOR.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC       ANCHOR (BY SIMILARITY).
CC   -!- PTM: BINDS FOUR HEME GROUPS PER MOLECULE.
CC   -!- SIMILARITY: HIGH, WITH OTHER PHOTOSYNTHETIC REACTION CENTER
CC       CYTOCHROME C SUBUNITS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB011811; BAA32742.1; -.
DR   InterPro; IPR000345; -.
DR   PROSITE; PS00190; CYTOCHROME_C; 4.
KW   Electron transport; Photosynthesis; Reaction center; Heme;
KW   Membrane; Lipoprotein; Duplication; Signal.
FT   SIGNAL       1     22       BY SIMILARITY.
FT   CHAIN       23    383       PHOTOSYNTHETIC REACTION CENTER CYTOCHROME
FT                               C SUBUNIT.
FT   LIPID        23     23       N-ACYL DIGLYCERIDE (BY SIMILARITY).
FT   BINDING     107    107       HEME 1 (COVALENT) (BY SIMILARITY).
FT   BINDING     110    110       HEME 1 (COVALENT) (BY SIMILARITY).
FT   METAL       111    111       IRON 1 (HEME AXIAL LIGAND)
FT                               (BY SIMILARITY).
FT   BINDING     152    152       HEME 2 (COVALENT) (BY SIMILARITY).
FT   BINDING     155    155       HEME 2 (COVALENT) (BY SIMILARITY).
FT   METAL       156    156       IRON 2 (HEME AXIAL LIGAND)
FT                               (BY SIMILARITY).
FT   BINDING     247    247       HEME 3 (COVALENT) (BY SIMILARITY).
FT   BINDING     250    250       HEME 3 (COVALENT) (BY SIMILARITY).
FT   METAL       251    251       IRON 3 (HEME AXIAL LIGAND)
FT                               (BY SIMILARITY).
FT   BINDING     307    307       HEME 4 (COVALENT) (BY SIMILARITY).
FT   BINDING     310    310       HEME 4 (COVALENT) (BY SIMILARITY).
FT   METAL       311    311       IRON 4 (HEME AXIAL LIGAND)
FT                               (BY SIMILARITY).
SQ   SEQUENCE   383 AA;  41522 MW;  96BCD91FF1B9AE7E CRC64;


  Query Match             40.7%;  Score 46;  DB 1;  Length 383;
  Best Local Similarity   42.1%;  Pred. No. 6.3;
  Matches    8;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPVGR 19
           ||    |||| | :| : :
Db     263 TPQRTTAWYAIRHVRDINQ 281


RESULT   6
YAAJ_ECOLI
ID   YAAJ_ECOLI     STANDARD;      PRT;   476 AA.
AC   P30143;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 51.7 KDA PROTEIN IN THRC-TALB INTERGENIC REGION (ORF8).
GN   YAAJ.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92334977; PubMed=1630901;
RA   Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA   Isono K., Mizobuchi K., Nakata A.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of
RT   the 0-2.4 min region.";
RL   Nucleic Acids Res. 20:3305-3308(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
```

```
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:ALANINE SYMPORTER FAMILY
CC       (SAF). STRONG, TO H.INFLUENZAE HI0183.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10483; -; NOT_ANNOTATED_CDS.
DR   EMBL; AE000111; AAC73118.1; -.
DR   EcoGene; EG11555; yaaJ.
DR   InterPro; IPR001463; -.
DR   Pfam; PF01235; Na_Ala_symp; 1.
DR   PRINTS; PR00175; NAALASMPORT.
DR   PROSITE; PS00873; NA_ALANINE_SYMP; 1.
KW   Hypothetical protein; Transmembrane; Inner membrane; Transport;
KW   Symport.
FT   TRANSMEM      4     24       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    141    161       POTENTIAL.
FT   TRANSMEM    174    194       POTENTIAL.
FT   TRANSMEM    207    227       POTENTIAL.
FT   TRANSMEM    233    253       POTENTIAL.
FT   TRANSMEM    300    320       POTENTIAL.
FT   TRANSMEM    351    371       POTENTIAL.
FT   TRANSMEM    391    411       POTENTIAL.
FT   TRANSMEM    414    434       POTENTIAL.
SQ   SEQUENCE   476 AA;  51662 MW;  2F6EB2E12E126E63 CRC64;


  Query Match             38.9%;  Score 44;  DB 1;  Length 476;
  Best Local Similarity   44.4%;  Pred. No. 16;
  Matches    8;  Conservative    3;  Mismatches    1;  Indels    6;  Gaps    1;

Qy       3 DIN------PAWYASRGI 14
           |:|      |||| :||:
Db     120 DVNGQFRGGPAWYMARGL 137


RESULT   7
ARP_YEAST
ID   ARP_YEAST      STANDARD;      PRT;   719 AA.
AC   P32770; Q12228;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ARP PROTEIN.
GN   ARP1 OR ARP OR YDL167C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AH22;
RX   MEDLINE=93247548; PubMed=8483449;
RA   Wehner E.P., Rao E., Brendel M.;
RT   "Molecular structure and genetic regulation of SFA, a gene
RT   responsible for resistance to formaldehyde in Saccharomyces
RT   cerevisiae, and characterization of its protein product.";
RL   Mol. Gen. Genet. 237:351-358(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RA   Pohl T.M.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 2 RANBP-TYPE ZINC FINGERS.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68020; CAA48159.1; -.
DR   EMBL; Z67750; CAA91579.1; -.
DR   EMBL; Z74215; CAA98741.1; -.
DR   PIR; S31139; S31139.
DR   HSSP; P04170; 6RXN.
DR   SGD; S0002326; NRP1.
DR   InterPro; IPR000504; -.
DR   InterPro; IPR001876; -.
DR   Pfam; PF00076; rrm; 1.
DR   Pfam; PF00641; zf-RanBP; 2.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW   Nuclear protein; Zinc-finger; RNA-binding; Repeat.
FT   DOMAIN      226    322       RNA-BINDING (RRM).
FT   DOMAIN      490    564       ASN-RICH.
FT   CONFLICT    493    493       I -> N (IN REF. 1).
SQ   SEQUENCE   719 AA;  79299 MW;  ADA9BC09FD582669 CRC64;


  Query Match             38.9%;  Score 44;  DB 1;  Length 719;
  Best Local Similarity   50.0%;  Pred. No. 24;
  Matches    6;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       7 AWYASRGIRPVG 18
             :|:   |:||||
Db     244 SWFTQYGVRPVG 255


RESULT   8
VG12_BPB03
ID   VG12_BPB03     STANDARD;      PRT;   860 AA.
AC   Q37893;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PRE-NECK APPENDAGE PROTEIN (LATE PROTEIN GP12).
GN   12.
OS   Bacteriophage B103.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.
OX   NCBI_TaxID=10778;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98019084; PubMed=9358052;
RA   Pecenkova T., Benes V., Paces J., Vlcek C., Paces V.;
RT   "Bacteriophage B103: complete DNA sequence of its genome and
RT   relationship to other Bacillus phages.";
RL   Gene 199:157-163(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X99260; CAA67660.1; -.
KW   Late protein.
SQ   SEQUENCE   860 AA;  91741 MW;  AC5B38F4BC65ECB0 CRC64;


  Query Match             38.5%;  Score 43.5;  DB 1;  Length 860;
```

```
  Best Local Similarity   50.0%;  Pred. No. 35;
  Matches   10;  Conservative    2;  Mismatches    3;  Indels    5;  Gaps    1;

Qy       2 PDINPAW-----YASRGIRP 16
           | |||::     | ||| ||
Db     780 PKINPSYNPKLEYLSRGERP 799


RESULT   9
OPS1_LIMPO
ID   OPS1_LIMPO     STANDARD;      PRT;   376 AA.
AC   P35360;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   LATERAL EYE OPSIN.
OS   Limulus polyphemus (Atlantic horseshoe crab).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC   Limulidae; Limulus.
OX   NCBI_TaxID=6850;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lateral eye;
RX   MEDLINE=93317641; PubMed=8327495;
RA   Smith W.C., Price D.A., Greenberg R.M., Battelle B.-A.;
RT   "Opsins from the lateral eyes and ocelli of the horseshoe crab,
RT   Limulus polyphemus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:6150-6154(1993).
CC   -!- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC       MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC       LINKED TO CIS-RETINAL.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: LATERAL EYE.
CC   -!- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC       BE PHOSPHORYLATED (BY SIMILARITY).
CC   -!- MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMA AT 520 NM.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       OPSIN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L03791; AAA28273.1; -.
DR   EMBL; L03781; AAA02498.1; -.
DR   PIR; B48197; B48197.
DR   GCRDb; GCR_0584; -.
DR   InterPro; IPR000276; -.
DR   InterPro; IPR001391; -.
DR   InterPro; IPR001760; -.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PRINTS; PR00238; OPSIN.
DR   PRINTS; PR00578; OPSINLTRLEYE.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
DR   PROSITE; PS00238; OPSIN; 1.
KW   Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW   Phosphorylation; G-protein coupled receptor.
FT   DOMAIN        1     46       EXTRACELLULAR.
FT   TRANSMEM     47     71       1 (POTENTIAL).
FT   DOMAIN       72     83       CYTOPLASMIC.
FT   TRANSMEM     84    108       2 (POTENTIAL).
FT   DOMAIN      109    123       EXTRACELLULAR.
FT   TRANSMEM    124    143       3 (POTENTIAL).
FT   DOMAIN      144    162       CYTOPLASMIC.
FT   TRANSMEM    163    186       4 (POTENTIAL).
FT   DOMAIN      187    210       EXTRACELLULAR.
FT   TRANSMEM    211    238       5 (POTENTIAL).
FT   DOMAIN      239    274       CYTOPLASMIC.
FT   TRANSMEM    275    298       6 (POTENTIAL).
FT   DOMAIN      299    306       EXTRACELLULAR.
FT   TRANSMEM    307    331       7 (POTENTIAL).
FT   DOMAIN      332    376       CYTOPLASMIC.
FT   DISULFID    120    197       BY SIMILARITY.
FT   BINDING     318    318       RETINAL CHROMOPHORE (BY SIMILARITY).
FT   CARBOHYD     17     17       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    193    193       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   376 AA;  42139 MW;  CCE401766AB06F26 CRC64;


  Query Match             38.1%;  Score 43;  DB 1;  Length 376;
  Best Local Similarity   46.2%;  Pred. No. 18;
  Matches    6;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYASRGI 14
           | :|| ||:  |:
Db      40 PPMNPLWYSILGV 52


RESULT  10
OPS2_LIMPO
ID   OPS2_LIMPO     STANDARD;      PRT;   376 AA.
AC   P35361;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   OCELLAR OPSIN.
OS   Limulus polyphemus (Atlantic horseshoe crab).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC   Limulidae; Limulus.
OX   NCBI_TaxID=6850;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Median ocelli;
RX   MEDLINE=93317641; PubMed=8327495;
RA   Smith W.C., Price D.A., Greenberg R.M., Battelle B.-A.;
RT   "Opsins from the lateral eyes and ocelli of the horseshoe crab,
RT   Limulus polyphemus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:6150-6154(1993).
CC   -!- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC       MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC       LINKED TO CIS-RETINAL.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: OCELLAR CELLS; MEDIAN OCELLI.
CC   -!- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC       BE PHOSPHORYLATED (BY SIMILARITY).
CC   -!- MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMA AT 530 NM.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       OPSIN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L03792; AAA28274.1; -.
DR   EMBL; L03782; AAA02499.1; -.
DR   PIR; A48197; A48197.
DR   GCRDb; GCR_0585; -.
DR   InterPro; IPR000276; -.
DR   InterPro; IPR001391; -.
DR   InterPro; IPR001760; -.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PRINTS; PR00238; OPSIN.
DR   PRINTS; PR00578; OPSINLTRLEYE.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
```

```
DR   PROSITE; PS00238; OPSIN; 1.
KW   Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW   Phosphorylation; G-protein coupled receptor.
FT   DOMAIN        1     46       EXTRACELLULAR.
FT   TRANSMEM     47     71       1 (POTENTIAL).
FT   DOMAIN       72     83       CYTOPLASMIC.
FT   TRANSMEM     84    108       2 (POTENTIAL).
FT   DOMAIN      109    123       EXTRACELLULAR.
FT   TRANSMEM    124    143       3 (POTENTIAL).
FT   DOMAIN      144    162       CYTOPLASMIC.
FT   TRANSMEM    163    186       4 (POTENTIAL).
FT   DOMAIN      187    210       EXTRACELLULAR.
FT   TRANSMEM    211    238       5 (POTENTIAL).
FT   DOMAIN      239    274       CYTOPLASMIC.
FT   TRANSMEM    275    298       6 (POTENTIAL).
FT   DOMAIN      299    306       EXTRACELLULAR.
FT   TRANSMEM    307    331       7 (POTENTIAL).
FT   DOMAIN      332    376       CYTOPLASMIC.
FT   DISULFID    120    197       BY SIMILARITY.
FT   BINDING     318    318       RETINAL CHROMOPHORE (BY SIMILARITY).
FT   CARBOHYD     17     17       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    193    193       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   376 AA;  42111 MW;  FA9647C40531CBF8 CRC64;


  Query Match             38.1%;  Score 43;  DB 1;  Length 376;
  Best Local Similarity   46.2%;  Pred. No. 18;
  Matches    6;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 PDINPAWYASRGI 14
          | :|| ||:  |:
Db     40 PPMNPLWYSILGV 52


RESULT  11
LYTB_MYCLE
ID   LYTB_MYCLE     STANDARD;      PRT;   332 AA.
AC   Q9X781;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   LYTB PROTEIN HOMOLOG.
GN   LYTB OR MLCB1222.06C.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Simmonds M.N., Badcock K., James K.D., Parkhill J., Barrell B.G.,
RA   Rajandream M.A.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE LYTB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL049491; CAB39812.1; ALT_INIT.
SQ   SEQUENCE   332 AA;  36225 MW;  F2A82A06EEC35E13 CRC64;


  Query Match             37.2%;  Score 42;  DB 1;  Length 332;
  Best Local Similarity   56.2%;  Pred. No. 23;
  Matches    9;  Conservative    2;  Mismatches    3;  Indels    2;  Gaps    1;

Qy      3 DINPAWYASRGIRPVG 18
          ||:||| |  |:  ||
Db    270 DIDPAWLA--GVTTVG 283


RESULT  12
YAIW_ECOLI
ID   YAIW_ECOLI     STANDARD;      PRT;   364 AA.
AC   P77562;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 40.4 KDA PROTEIN IN SBMA-DDLA INTERGENIC REGION.
GN   YAIW.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA   Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA   Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000144; AAC73481.1; -.
DR   EMBL; U73857; AAB18101.1; -.
DR   EcoGene; EG13608; yaiW.
KW   Hypothetical protein.
SQ   SEQUENCE   364 AA;  40414 MW;  7170FAF05C4E5973 CRC64;


  Query Match             37.2%;  Score 42;  DB 1;  Length 364;
  Best Local Similarity   70.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 DINPAWYASR 12
          | |  |||||
Db    233 DFNAGWYASR 242


RESULT  13
CHOD_STRSQ
ID   CHOD_STRSQ     STANDARD;      PRT;   546 AA.
AC   P12676;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CHOLESTEROL OXIDASE PRECURSOR (EC 1.1.3.6) (CHOD).
GN   CHOA.
OS   Streptomyces sp. (strain SA-COO).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1931;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89123081; PubMed=2914858;
RA   Ishizaki T., Hirayama N., Shinkawa H., Nimi O., Murooka Y.;
```

```
RT   "Nucleotide sequence of the gene for cholesterol oxidase from a
RT   Streptomyces sp.";
RL   J. Bacteriol. 171:596-601(1989).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS), AND MUTAGENESIS.
RX   MEDLINE=99211873; PubMed=10194345;
RA   Yue Q.K., Kass I.J., Sampson N.S., Vrielink A.;
RT   "Crystal structure determination of cholesterol oxidase from
RT   Streptomyces and structural characterization of key active site
RT   mutants.";
RL   Biochemistry 38:4277-4286(1999).
CC   -!- CATALYTIC ACTIVITY: CHOLESTEROL + O(2) = CHOLEST-4-EN-3-ONE +
CC       H(2)O(2).
CC   -!- COFACTOR: FAD.
CC   -!- PATHWAY: CHOLESTEROL METABOLISM.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31939; AAA26719.1; -.
DR   PIR; A32260; A32260.
DR   PDB; 1B4V; 06-JAN-99.
DR   PDB; 1B8S; 09-FEB-99.
DR   PDB; 1CBO; 10-MAR-99.
DR   PDB; 1CC2; 11-MAR-99.
DR   InterPro; IPR000172; -.
DR   InterPro; IPR001167; -.
DR   Pfam; PF01319; CHOD; 1.
DR   PROSITE; PS00623; GMC_OXRED_1; 1.
DR   PROSITE; PS00624; GMC_OXRED_2; FALSE_NEG.
KW   Oxidoreductase; Signal; Flavoprotein; FAD; 3D-structure.
FT   SIGNAL       1     42
FT   CHAIN       43    546       CHOLESTEROL OXIDASE.
FT   NP_BIND     54     70       FAD (ADP PART) (POTENTIAL).
FT   ACT_SITE   398    398       PROTON ACCEPTOR.
FT   ACT_SITE   484    484
FT   MUTAGEN    398    398       E->Q: REDUCED ACTIVITY.
FT   MUTAGEN    484    484       H->N,Q: REDUCED ACTIVITY.
SQ   SEQUENCE   546 AA;  58993 MW;  EF22A1FE5EA68D21 CRC64;


  Query Match             37.2%;  Score 42;  DB 1;  Length 546;
  Best Local Similarity   36.8%;  Pred. No. 38;
  Matches    7;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2 PDINPAWYASRGIRPVGRF 20
           ||   :|: :|   |:| |
Db      98 PDKRSSWFKNRTEAPLGSF 116


RESULT  14
POL_MLVRK
ID   POL_MLVRK      STANDARD;      PRT;   581 AA.
AC   P31795;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   POL POLYPROTEIN [CONTAINS: PROTEASE (EC 3.4.23.-); REVERSE
DE   TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)] (FRAGMENT).
GN   POL.
OS   Radiation murine leukemia virus (strain Kaplan).
OC   Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
OX   NCBI_TaxID=31689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92333703; PubMed=1629969;
RA   Poliquin L., Bergeron D., Fortier J.L., Paquette Y., Bergeron R.,
RA   Rassart E.;
RT   "Determinants of thymotropism in Kaplan radiation leukemia virus and
RT   nucleotide sequence of its envelope region.";
RL   J. Virol. 66:5141-5146(1992).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
CC       POLYPROTEIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A2; ALSO KNOWN AS THE
CC       RETROPEPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M93052; AAA46525.1; -.
DR   PIR; A42743; A42743.
DR   MEROPS; A02.008; -.
DR   InterPro; IPR001037; -.
DR   InterPro; IPR001584; -.
DR   InterPro; IPR001969; -.
DR   InterPro; IPR002156; -.
DR   Pfam; PF00552; integrase; 1.
DR   Pfam; PF00075; rnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   PROSITE; PS00141; ASP_PROTEASE; PARTIAL.
KW   Hydrolase; Transferase; RNA-directed DNA polymerase;
KW   Aspartyl protease; Endonuclease; Polyprotein.
FT   NON_TER      1      1
SQ   SEQUENCE   581 AA;  65157 MW;  8D7A38694C8E036E CRC64;


  Query Match             37.2%;  Score 42;  DB 1;  Length 581;
  Best Local Similarity   41.2%;  Pred. No. 41;
  Matches    7;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPV 17
           || | |:|   |   |:
Db     556 TPPIRPSWRVQRSQNPL 572


RESULT  15
POL_MLVAK
ID   POL_MLVAK      STANDARD;      PRT;   843 AA.
AC   P03357;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE   ENDONUCLEASE] (FRAGMENT).
GN   POL.
OS   AKR murine leukemia virus.
OC   Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
OX   NCBI_TaxID=11789;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83090450; PubMed=6294621;
RA   Herr W., Corbin V., Gilbert W.;
RT   "Nucleotide sequence of the 3' half of AKV.";
RL   Nucleic Acids Res. 10:6931-6944(1982).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
DR   PIR; A03958; GNVWK.
DR   HSSP; P03355; 1MML.
DR   InterPro; IPR001037; -.
DR   InterPro; IPR001584; -.
DR   InterPro; IPR002156; -.
DR   Pfam; PF00552; integrase; 1.
DR   Pfam; PF00075; rnaseH; 1.
```

```
DR   Pfam; PF00665; rve; 1.
KW   Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW   Endonuclease; Polyprotein.
FT   NON_TER       1      1
SQ   SEQUENCE   843 AA;  94111 MW;  274C97731EE16C6A CRC64;


  Query Match             37.2%;  Score 42;  DB 1;  Length 843;
  Best Local Similarity   41.2%;  Pred. No. 59;
  Matches    7;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPV 17
           || | |:|   |   |:
Db     818 TPPIKPSWRVQRSQNPL 834


Search completed: April 17, 2001, 15:48:52
Job time: 540 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:39:48 ; Search time 61.54 Seconds
(without alignments)
6.243 Million cell updates/sec

Title: US-09-446-543A-64
Perfect score: 113
Sequence: 1 TPDINPAWYASRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	113	100.0	20	3	US-09-105-678A-46	Sequence 46, Appl
2	113	100.0	21	3	US-09-105-678A-47	Sequence 47, Appl
3	113	100.0	22	3	US-09-105-678A-48	Sequence 48, Appl
4	113	100.0	31	3	US-09-105-678A-9	Sequence 9, Appli
5	113	100.0	31	3	US-09-105-678A-43	Sequence 43, Appl
6	113	100.0	32	3	US-09-105-678A-44	Sequence 44, Appl
7	113	100.0	33	3	US-09-105-678A-45	Sequence 45, Appl
8	109	96.5	20	3	US-09-105-678A-34	Sequence 34, Appl
9	109	96.5	21	3	US-09-105-678A-35	Sequence 35, Appl
10	109	96.5	22	3	US-09-105-678A-36	Sequence 36, Appl
11	109	96.5	31	3	US-09-105-678A-7	Sequence 7, Appli
12	109	96.5	31	3	US-09-105-678A-31	Sequence 31, Appl
13	109	96.5	32	3	US-09-105-678A-32	Sequence 32, Appl
14	109	96.5	33	3	US-09-105-678A-33	Sequence 33, Appl
15	105	92.9	20	3	US-09-105-678A-40	Sequence 40, Appl
16	105	92.9	21	3	US-09-105-678A-28	Sequence 28, Appl
17	105	92.9	21	3	US-09-105-678A-41	Sequence 41, Appl
18	105	92.9	22	3	US-09-105-678A-42	Sequence 42, Appl
19	105	92.9	31	3	US-09-105-678A-8	Sequence 8, Appli
20	105	92.9	31	3	US-09-105-678A-37	Sequence 37, Appl
21	105	92.9	31	4	US-09-172-353-4	Sequence 4, Appli
22	105	92.9	32	3	US-09-105-678A-38	Sequence 38, Appl
23	105	92.9	33	3	US-09-105-678A-39	Sequence 39, Appl
24	103	91.2	19	3	US-09-105-678A-30	Sequence 30, Appl
25	98	86.7	29	3	US-09-105-678A-29	Sequence 29, Appl
26	44	38.9	349	1	US-08-118-270-71	Sequence 71, Appl
27	44	38.9	349	5	PCT-US93-08528-71	Sequence 71, Appl
28	42	37.2	547	1	US-08-083-948-8	Sequence 8, Appli
29	42	37.2	547	1	US-08-393-785-8	Sequence 8, Appli
30	42	37.2	547	1	US-08-475-694-8	Sequence 8, Appli
31	42	37.2	547	1	US-08-712-057-8	Sequence 8, Appli
32	40	35.4	433	1	US-08-417-492-2	Sequence 2, Appli
33	40	35.4	555	2	US-08-982-232-7	Sequence 7, Appli
34	40	35.4	555	2	US-08-982-232-14	Sequence 14, Appl
35	39.5	35.0	831	1	US-08-073-384C-5	Sequence 5, Appli
36	39.5	35.0	831	1	US-08-254-359A-5	Sequence 5, Appli
37	39.5	35.0	831	1	US-08-483-043-5	Sequence 5, Appli
38	39.5	35.0	831	1	US-08-481-238-5	Sequence 5, Appli
39	39.5	35.0	831	2	US-08-471-066B-5	Sequence 5, Appli
40	39.5	35.0	831	2	US-08-484-956-5	Sequence 5, Appli
41	39.5	35.0	831	2	US-08-757-653-5	Sequence 5, Appli
42	39.5	35.0	831	2	US-08-599-491-5	Sequence 5, Appli
43	39.5	35.0	831	2	US-08-756-386-5	Sequence 5, Appli
44	39.5	35.0	831	2	US-08-823-516-5	Sequence 5, Appli
45	39.5	35.0	831	3	US-08-682-853A-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT 1
US-09-105-678A-46
; Sequence 46, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-46


  Query Match           100.0%;  Score 113;  DB 3;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 1e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 TPDINPAWYASRGIRPVGRF 20


RESULT   2
US-09-105-678A-47
; Sequence 47, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-47


  Query Match             100.0%;  Score 113;  DB 3;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 TPDINPAWYASRGIRPVGRF 20


RESULT   3
US-09-105-678A-48
; Sequence 48, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-48


  Query Match             100.0%;  Score 113;  DB 3;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 TPDINPAWYASRGIRPVGRF 20


RESULT   4
US-09-105-678A-9
; Sequence 9, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
```

```
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-9


  Query Match           100.0%;  Score 113;  DB 3;  Length 31;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYASRGIRPVGRF 31


RESULT   5
US-09-105-678A-43
; Sequence 43, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  43:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-43


  Query Match           100.0%;  Score 113;  DB 3;  Length 31;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYASRGIRPVGRF 31


RESULT   6
US-09-105-678A-44
; Sequence 44, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-44


  Query Match           100.0%;  Score 113;  DB 3;  Length 32;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYASRGIRPVGRF 31


RESULT   7
US-09-105-678A-45
; Sequence 45, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-45


  Query Match             100.0%;  Score 113;  DB 3;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 TPDINPAWYASRGIRPVGRF 31


RESULT   8
US-09-105-678A-34
; Sequence 34, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-34


  Query Match             96.5%;  Score 109;  DB 3;  Length 20;
  Best Local Similarity   95.0%;  Pred. No. 4.4e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          |||||||||| |||||||||
Db      1 TPDINPAWYAGRGIRPVGRF 20


RESULT   9
US-09-105-678A-35
; Sequence 35, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-35


  Query Match             96.5%;  Score 109;  DB 3;  Length 21;
  Best Local Similarity   95.0%;  Pred. No. 4.6e-11;
```

```
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          |||||||||| |||||||||
Db      1 TPDINPAWYAGRGIRPVGRF 20


RESULT  10
US-09-105-678A-36
; Sequence 36, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-36


  Query Match           96.5%;  Score 109;  DB 3;  Length 22;
  Best Local Similarity  95.0%;  Pred. No. 4.9e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          |||||||||| |||||||||
Db      1 TPDINPAWYAGRGIRPVGRF 20


RESULT  11
US-09-105-678A-7
; Sequence 7, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-7


  Query Match           96.5%;  Score 109;  DB 3;  Length 31;
  Best Local Similarity  95.0%;  Pred. No. 7.2e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          |||||||||| |||||||||
Db     12 TPDINPAWYAGRGIRPVGRF 31


RESULT  12
US-09-105-678A-31
; Sequence 31, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-31


  Query Match             96.5%;  Score 109;  DB 3;  Length 31;
  Best Local Similarity   95.0%;  Pred. No. 7.2e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          |||||||||| |||||||||
Db     12 TPDINPAWYAGRGIRPVGRF 31


RESULT  13
US-09-105-678A-32
; Sequence 32, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-32


  Query Match             96.5%;  Score 109;  DB 3;  Length 32;
  Best Local Similarity   95.0%;  Pred. No. 7.4e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          |||||||||| |||||||||
Db     12 TPDINPAWYAGRGIRPVGRF 31


RESULT  14
US-09-105-678A-33
; Sequence 33, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-33


  Query Match             96.5%;  Score 109;  DB 3;  Length 33;
  Best Local Similarity   95.0%;  Pred. No. 7.7e-11;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          |||||||||| |||||||||
Db     12 TPDINPAWYAGRGIRPVGRF 31


RESULT  15
US-09-105-678A-40
; Sequence 40, Application US/09105678A
; Patent No. 6103882
;  GENERAL INFORMATION:
;    APPLICANT:  Suenaga, Masato
;    APPLICANT:  Moriya, Takeo
;    APPLICANT:  Tanaka, Yoko
;    APPLICANT:  Nishimura, Osamu
;    TITLE OF INVENTION:  METHOD OF PRODUCING A 19P2 LIGAND
```

```
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/105,678A
;      FILING DATE:  26-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 172118/1997
;      FILING DATE:  27-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conlin, David G.
;      REGISTRATION NUMBER:  27,026
;      REFERENCE/DOCKET NUMBER:  48466-342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-105-678A-40


  Query Match             92.9%;  Score 105;  DB 3;  Length 20;
  Best Local Similarity   90.0%;  Pred. No. 1.9e-10;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPVGRF 20
           |||||||||  |||||||||
Db       1 TPDINPAWYTGRGIRPVGRF 20


Search completed: April 17, 2001, 15:39:48
Job time: 317 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:38:42 ; Search time 116.94 Seconds
(without alignments)
9.777 Million cell updates/sec

Title: US-09-446-543A-64
Perfect score: 113
Sequence: 1 TPDINPAWYASRGIRPVGRF 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	113	100.0	20	18	W31394	Human type G prote
2	113	100.0	20	20	W97236	Human type ligand
3	113	100.0	20	21	B10365	Human oxytocin sec
4	113	100.0	20	21	Y49294	19P2 ligand peptid
5	113	100.0	21	18	W31395	Human type G prote
6	113	100.0	21	21	B10366	Human oxytocin sec
7	113	100.0	22	18	W31396	Human type G prote
8	113	100.0	22	21	B10367	Human oxytocin sec
9	113	100.0	31	18	W31391	Human type G prote
10	113	100.0	31	20	W97235	Human type ligand
11	113	100.0	31	20	W87615	Human 19P2 ligand.
12	113	100.0	31	21	B10362	Human oxytocin sec
13	113	100.0	31	21	Y49291	19P2 ligand peptid
14	113	100.0	32	18	W31392	Human type G prote
15	113	100.0	32	21	B10363	Human oxytocin sec
16	113	100.0	33	18	W31393	Human type G prote
17	113	100.0	33	21	B10364	Human oxytocin sec
18	113	100.0	87	18	W31390	Human type G prote
19	113	100.0	87	20	W97226	Human type ligand
20	113	100.0	87	21	B10361	Human oxytocin sec
21	109	96.5	20	18	W31374	Bovine G protein-c
22	109	96.5	20	20	W97232	Bovine pituitary-d
23	109	96.5	20	20	W95191	Bovine pituitary-d
24	109	96.5	20	21	B10350	Bovine oxytocin se
25	109	96.5	20	21	Y49301	19P2 ligand peptid
26	109	96.5	21	18	W31375	Bovine G protein-c
27	109	96.5	21	20	W95192	Bovine pituitary-d
28	109	96.5	21	21	B10351	Bovine oxytocin se
29	109	96.5	22	18	W31376	Bovine G protein-c
30	109	96.5	22	20	W95193	Bovine pituitary-d
31	109	96.5	22	21	B10352	Bovine oxytocin se
32	109	96.5	31	18	W31371	Bovine G protein-c
33	109	96.5	31	20	W97218	Bovine pituitary-d
34	109	96.5	31	20	W87613	Bovine 19P2 ligand
35	109	96.5	31	20	W95188	Bovine pituitary-d
36	109	96.5	31	21	B10347	Bovine oxytocin se
37	109	96.5	31	21	Y49290	19P2 ligand peptid
38	109	96.5	31	21	Y49298	19P2 ligand peptid
39	109	96.5	32	18	W31372	Bovine G protein-c
40	109	96.5	32	20	W95189	Bovine pituitary-d
41	109	96.5	32	21	B10348	Bovine oxytocin se
42	109	96.5	33	18	W31373	Bovine G protein-c
43	109	96.5	33	20	W95190	Bovine pituitary-d
44	109	96.5	33	21	B10349	Bovine oxytocin se
45	109	96.5	33	21	Y49297	19P2 ligand peptid

ALIGNMENTS

RESULT 1
W31394
ID W31394 standard; Peptide; 20 AA.
XX
AC W31394;
XX
DT 06-APR-1998 (first entry)
XX
DE Human type G protein-coupled receptor ligand fragment 4.
XX
KW G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
OS Homo sapiens.
XX
PN WO9724436-A2.
XX
PD 10-JUL-1997.
XX
PF 26-DEC-1996; 96WO-JP03821.
XX
PR 18-SEP-1996; 96JP-0246573.
PR 28-DEC-1995; 95JP-0343371.
PR 15-MAR-1996; 96JP-0059419.
PR 12-AUG-1996; 96JP-0211805.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI Kawamata Y, Kitada C;
XX
DR WPI; 1997-363672/33.
DR N-PSDB; V02431.

XX
PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2; Page 185; 258pp; English.
XX
CC This sequence represents a peptide fragment from a novel human type
CC ligand polypeptide corresponding to amino acid residues 34 to 53 of the
CC sequence represented in W31390 and is used in an assay to monitor ligand
CC binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a prophylactic
CC or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC compounds which are capable of altering the binding activity of the
CC ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 113; DB 18; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYASRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawyasrgirpvgrf 20

RESULT 2
W97236
ID W97236 standard; peptide; 20 AA.
XX
AC W97236;
XX
DT 06-MAY-1999 (first entry)
XX
DE Human type ligand polypeptide fragment.
XX
KW Rat type ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypoovarianism; gonecyst cacogenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmeniopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castilo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.
XX
OS Homo sapiens.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI; 1999-105614/09.
XX
PT Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Claim 3; Page 166; 241pp; English.
XX
CC The present sequence represents a human type ligand fragment. It
CC is used in the course of the invention. The specification describes
CC an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypoovarianism, gonecyst cacogenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can by used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmeniopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castilo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, irruption mole, abortion, unthrifty fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 113; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYASRGIRPVGRF 20
||||||||||||||||||||
Db 1 tpdinpawyasrgirpvgrf 20

RESULT 3
B10365
ID B10365 standard; peptide; 20 AA.
XX
AC B10365;
XX
DT 24-NOV-2000 (first entry)
XX
DE Human oxytocin secretion promoting peptide SEQ ID NO: 35.
XX
KW Human; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX
OS Homo sapiens.
XX
PN WO200038704-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-JP07199.
XX
PR 25-DEC-1998; 98JP-0369585.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-452298/39.
XX
PT Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine -
XX

```
PS   Disclosure; Page 63; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a human peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   20 AA;


  Query Match            100.0%;  Score 113;  DB 21;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 3.5e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawyasrgirpvgrf 20


RESULT   4
Y49294
ID   Y49294 standard; peptide; 20 AA.
XX
AC   Y49294;
XX
DT   22-FEB-2000  (first entry)
XX
DE   19P2 ligand peptide fragment.
XX
KW   Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW   pituitary; regulatory mechanism; central nervous system; pancreatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   20
FT                   /note= "C-terminal amide"
XX
PN   WO9960112-A1.
XX
PD   25-NOV-1999.
XX
PF   20-MAY-1999;   99WO-JP02650.
XX
PR   21-MAY-1998;   98JP-0140293.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-039381/03.
XX
PT   New monoclonal antibodies,  useful in diagnosis, as drugs and  in
PT   studying diseases related to ligand abnormality  -
XX
PS   Disclosure; Page 26; 73pp; Japanese.
XX
CC   The invention provides a monoclonal antibody which has a specific
CC   reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC   derivative. The antibodies can be used in diagnosis or to treat or
CC   prevent diseases associated with abnormality in the pituitary function
CC   regulatory mechanism (e.g. promotion of prolactin secretion), central
CC   nervous regulatory mechanism, and pancreatic function regulatory
CC   mechanism. The antibody-based immunoassay can also be applied in
CC   clarifying the physiological functions of the ligand and its derivative.
CC   Sequences Y49290-302 represent peptide fragments of the 19P2 ligand.
XX
SQ   Sequence   20 AA;


  Query Match            100.0%;  Score 113;  DB 21;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 3.5e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawyasrgirpvgrf 20


RESULT   5
W31395
ID   W31395 standard; Peptide; 21 AA.
XX
AC   W31395;
XX
DT   06-APR-1998  (first entry)
XX
DE   Human type G protein-coupled receptor ligand fragment 5.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Homo sapiens.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02432.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 186; 258pp; English.
XX
CC   This sequence represents a peptide fragment from a novel human type
CC   ligand polypeptide corresponding to amino acid residues 34 to 54 of the
CC   sequence represented in W31390 and is used in an assay to monitor ligand
CC   binding to the G protein-coupled receptor protein. Pharmaceutical
CC   compositions containing this ligand may be used as a pituitary function
CC   modulator, a central nervous system modulator or a pancreatic function
CC   modulator. This ligand could have specific applications as a prophylactic
CC   or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC   disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC   growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC   hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC   amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC   compounds which are capable of altering the binding activity of the
CC   ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ   Sequence   21 AA;
```

```
  Query Match             100.0%;  Score 113;  DB 18;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPVGRF 20
           ||||||||||||||||||||
Db       1 tpdinpawyasrgirpvgrf 20


RESULT   6
B10366
ID   B10366 standard; peptide; 21 AA.
XX
AC   B10366;
XX
DT   24-NOV-2000  (first entry)
XX
DE   Human oxytocin secretion promoting peptide SEQ ID NO: 36.
XX
KW   Human; oxytocin secretion promoter; G protein-coupled receptor protein;
KW   treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW   caesarean section; artificial fertilization; galactostasis; goat; pig;
KW   veterinary medicine; milk production.
XX
OS   Homo sapiens.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine  -
XX
PS   Disclosure; Page 63; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a human peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   21 AA;


  Query Match             100.0%;  Score 113;  DB 21;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPVGRF 20
           ||||||||||||||||||||
Db       1 tpdinpawyasrgirpvgrf 20


RESULT   7
W31396
```

```
ID   W31396 standard; Peptide; 22 AA.
XX
AC   W31396;
XX
DT   06-APR-1998  (first entry)
XX
DE   Human type G protein-coupled receptor ligand fragment 6.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Homo sapiens.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02433.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 186; 258pp; English.
XX
CC   This sequence represents a peptide fragment from a novel human type
CC   ligand polypeptide corresponding to amino acid residues 34 to 55 of the
CC   sequence represented in W31390 and is used in an assay to monitor ligand
CC   binding to the G protein-coupled receptor protein. Pharmaceutical
CC   compositions containing this ligand may be used as a pituitary function
CC   modulator, a central nervous system modulator or a pancreatic function
CC   modulator. This ligand could have specific applications as a prophylactic
CC   or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC   disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC   growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC   hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC   amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC   compounds which are capable of altering the binding activity of the
CC   ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ   Sequence   22 AA;


  Query Match             100.0%;  Score 113;  DB 18;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPDINPAWYASRGIRPVGRF 20
           ||||||||||||||||||||
Db       1 tpdinpawyasrgirpvgrf 20


RESULT   8
B10367
ID   B10367 standard; peptide; 22 AA.
XX
```

```
AC   B10367;
XX
DT   24-NOV-2000  (first entry)
XX
DE   Human oxytocin secretion promoting peptide SEQ ID NO: 37.
XX
KW   Human; oxytocin secretion promoter; G protein-coupled receptor protein;
KW   treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW   caesarean section; artificial fertilization; galactostasis; goat; pig;
KW   veterinary medicine; milk production.
XX
OS   Homo sapiens.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine -
XX
PS   Disclosure; Page 64; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a human peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   22 AA;


  Query Match             100.0%;  Score 113;  DB 21;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db      1 tpdinpawyasrgirpvgrf 20


RESULT   9
W31391
ID   W31391 standard; Peptide; 31 AA.
XX
AC   W31391;
XX
DT   06-APR-1998  (first entry)
XX
DE   Human type G protein-coupled receptor ligand fragment 1.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Homo sapiens.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02428.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 184; 258pp; English.
XX
CC   This sequence represents a peptide fragment from a novel human type
CC   ligand polypeptide corresponding to amino acid residues 23 to 53 of the
CC   sequence represented in W31390 and is used in an assay to monitor ligand
CC   binding to the G protein-coupled receptor protein. Pharmaceutical
CC   compositions containing this ligand may be used as a pituitary function
CC   modulator, a central nervous system modulator or a pancreatic function
CC   modulator. This ligand could have specific applications as a prophylactic
CC   or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC   disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC   growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC   hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC   cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC   rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC   amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC   spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC   osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC   compounds which are capable of altering the binding activity of the
CC   ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ   Sequence   31 AA;


  Query Match             100.0%;  Score 113;  DB 18;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 tpdinpawyasrgirpvgrf 31


RESULT  10
W97235
ID   W97235 standard; peptide; 31 AA.
XX
AC   W97235;
XX
DT   06-MAY-1999  (first entry)
XX
DE   Human type ligand polypeptide fragment.
XX
KW   Rat type ligand; modulation; prolactin secretion;
KW   G protein-coupled receptor; GPCR; hypoovarianism; gonecyst cacogenesis;
KW   menopausal syndrome; euthyroid; hypometabolism; lactation;
KW   pituitary adenomatosis; brain tumour; emmeniopathy; autoimmune disease;
KW   prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW   acromegaly; Chiari-Frommel syndrome; Argonz-del Castilo syndrome;
KW   Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW   contraceptive; placental function; choriocarcinoma; hydatid mole;
```

KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.
XX
OS Homo sapiens.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI; 1999-105614/09.
XX
PT Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
PS Claim 3; Page 159; 241pp; English.
XX
CC The present sequence represents a human type ligand fragment. It
CC is used in the course of the invention. The specification describes
CC an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypoovarianism, gonecyst cacogenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can by used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmeniopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castilo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, irruption mole, abortion, unthrifty fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 113; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYASRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawyasrgirpvgrf 31

RESULT 11
W87615
ID W87615 standard; Peptide; 31 AA.
XX
AC W87615;
XX
DT 29-MAR-1999 (first entry)
XX
DE Human 19P2 ligand.
XX
KW 19P2 ligand; G protein coupled receptor; pituitary;
KW prolactin releasing peptide; human; dementia; breast cancer;
KW therapy.
XX
OS Homo sapiens.
XX
PN EP887417-A2.
XX
PD 30-DEC-1998.
XX
PF 25-JUN-1998; 98EP-0111725.
XX
PR 27-JUN-1997; 97JP-0172118.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Moriya T, Nishimura O, Suenaga M, Tanaka Y;
XX
DR WPI; 1999-047884/05.
XX
PT Producing a 19P2 pituitary G protein receptor ligand - by cleavage
PT of a fusion protein, useful for preventing and treating dementia,
PT breast cancer, renal failure and autoimmune disease
XX
PS Claim 5; Page 35; 56pp; English.
XX
CC This is the amino acid sequence of the human pituitary G
CC protein-coupled receptor ligand 19P2L. A method suitable for
CC commercial high-level production of 19P2L comprises expressing
CC the ligand in host cells as a recombinant fusion protein e.g. with
CC human basic fibroblast growth factor (see V83796-97) that has
CC been modified to include an N-terminal cysteine residue. The
CC ligand is released from the fusion by cyanylation followed by
CC ammonolysis. 19P2L has prolactin secretion-stimulating and (at
CC high doses) prolactin secretion-inhibiting properties. It can be
CC used in the treatment and prevention of various diseases including:
CC senile dementia, cerebrovascular dementia, and dementia associated
CC with: genealogical disorders (e.g. Alzheimer's disease, Parkinson's
CC disease, Pick's disease, Huntington's disease), infectious diseases
CC (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or
CC toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism,
CC intoxication by drugs, metal and organic compounds), tumourigenic
CC diseases (e.g. brain tumour), traumatic diseases (e.g. chronic
CC subarachnoidal heamorrhage, and other types of dementia, depression,
CC hyperactive child syndrome (microencephalopathy) and disturbance of
CC consciousness. It is also useful for prevention and treatment of
CC diseases associated with prolactin hypo and hypersecretion
CC respectively, including: hyperprolactinaemia, pituitary adenoma,
CC breast cancer, infertility, impotence and autoimmune disease
CC (hypersecretion disorders), and seminal vesicle hypoplasia,
CC osteoporosis, menopausal syndrome and renal failure (hyposecretion
CC disorders). The 19P2 polypeptide/amide is also useful as a test
CC reagent for study of the prolactin secretory function or as a
CC lactogogue in mammalian farm animals.
XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 113; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYASRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawyasrgirpvgrf 31

RESULT 12
B10362
ID B10362 standard; peptide; 31 AA.
XX
AC B10362;
XX
DT 24-NOV-2000 (first entry)
XX
DE Human oxytocin secretion promoting peptide SEQ ID NO: 32.
XX
KW Human; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.

```
XX
OS   Homo sapiens.
XX
PN   WO200038704-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-JP07199.
XX
PR   25-DEC-1998;   98JP-0369585.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-452298/39.
XX
PT   Physiologically-active polypeptide recognized as ligand by G
PT   protein-coupled receptor protein, for promoting secretion of oxytocin,
PT   as drugs for diseases relating to oxytocin secretion and in veterinary
PT   medicine  -
XX
PS   Disclosure; Page 62; 72pp; Japanese.
XX
CC   This invention describes a novel oxytocin secretion-regulating agent
CC   which contains a ligand peptide or its salt for the G protein-coupled
CC   receptor protein. It is useful in the form of drugs for ameliorating,
CC   preventing and treating diseases relating to oxytocin secretion e.g.
CC   weak pains and atonic bleeding, before and after expulsion of placenta,
CC   uterine recovery failure, caesarean section, stoppage of artificial
CC   fertilization or galactostasis and is also applicable in veterinary
CC   medicine for promoting milk production in cow, goat and pig. This
CC   sequence represents a human peptide which acts as an oxytocin secretion
CC   promoter.
XX
SQ   Sequence   31 AA;


  Query Match             100.0%;  Score 113;  DB 21;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 tpdinpawyasrgirpvgrf 31


RESULT  13
Y49291
ID   Y49291 standard; peptide; 31 AA.
XX
AC   Y49291;
XX
DT   22-FEB-2000  (first entry)
XX
DE   19P2 ligand peptide fragment.
XX
KW   Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW   pituitary; regulatory mechanism; central nervous system; pancreatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   31
FT                   /note= "C-terminal amide"
XX
PN   WO9960112-A1.
XX
PD   25-NOV-1999.
XX
PF   20-MAY-1999;   99WO-JP02650.
XX
PR   21-MAY-1998;   98JP-0140293.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Matsumoto H,  Kitada C,  Hinuma S;
XX
DR   WPI; 2000-039381/03.
XX
PT   New monoclonal antibodies,  useful in diagnosis, as drugs and  in
PT   studying diseases related to ligand abnormality  -
XX
PS   Disclosure; Page 26; 73pp; Japanese.
XX
CC   The invention provides a monoclonal antibody which has a specific
CC   reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC   derivative. The antibodies can be used in diagnosis or to treat or
CC   prevent diseases associated with abnormality in the pituitary function
CC   regulatory mechanism (e.g. promotion of prolactin secretion), central
CC   nervous regulatory mechanism, and pancreatic function regulatory
CC   mechanism. The antibody-based immunoassay can also be applied in
CC   clarifying the physiological functions of the ligand and its derivative.
CC   Sequences Y49290-302 represent peptide fragments of the 19P2 ligand.
XX
SQ   Sequence   31 AA;


  Query Match             100.0%;  Score 113;  DB 21;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPDINPAWYASRGIRPVGRF 20
          ||||||||||||||||||||
Db     12 tpdinpawyasrgirpvgrf 31


RESULT  14
W31392
ID   W31392 standard; Peptide; 32 AA.
XX
AC   W31392;
XX
DT   06-APR-1998  (first entry)
XX
DE   Human type G protein-coupled receptor ligand fragment 2.
XX
KW   G protein-coupled receptor; ligand binding; pharmaceutical;
KW   modulator; pituitary; central nervous system; pancreas; prophylactic;
KW   therapeutic agent.
XX
OS   Homo sapiens.
XX
PN   WO9724436-A2.
XX
PD   10-JUL-1997.
XX
PF   26-DEC-1996;   96WO-JP03821.
XX
PR   18-SEP-1996;   96JP-0246573.
PR   28-DEC-1995;   95JP-0343371.
PR   15-MAR-1996;   96JP-0059419.
PR   12-AUG-1996;   96JP-0211805.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fujii R,  Fukusumi S,  Habata Y,  Hinuma S,  Hosoya M;
PI   Kawamata Y,  Kitada C;
XX
DR   WPI; 1997-363672/33.
DR   N-PSDB; V02429.
XX
PT   Ligand peptide for G protein-coupled receptor - acts by modulating
PT   function in the central nervous system, pancreas and pituitary gland
XX
PS   Claim 2; Page 185; 258pp; English.
```

XX
CC This sequence represents a peptide fragment from a novel human type
CC ligand polypeptide corresponding to amino acid residues 23 to 54 of the
CC sequence represented in W31390 and is used in an assay to monitor ligand
CC binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a prophylactic
CC or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
CC hypercholesterolaemia, hyperglyceridaemia, hyperprolactinaemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC amylotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC compounds which are capable of altering the binding activity of the
CC ligand affecting activation of the G protein-coupled receptor protein.
XX
SQ Sequence 32 AA;

Query Match 100.0%; Score 113; DB 18; Length 32;
Best Local Similarity 100.0%; Pred. No. 6.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYASRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawyasrgirpvgrf 31

RESULT 15
B10363
ID B10363 standard; peptide; 32 AA.
XX
AC B10363;
XX
DT 24-NOV-2000 (first entry)
XX
DE Human oxytocin secretion promoting peptide SEQ ID NO: 33.
XX
KW Human; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX
OS Homo sapiens.
XX
PN WO200038704-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-JP07199.
XX
PR 25-DEC-1998; 98JP-0369585.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI; 2000-452298/39.
XX
PT Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine -
XX
PS Disclosure; Page 62; 72pp; Japanese.
XX
CC This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a human peptide which acts as an oxytocin secretion
CC promoter.
XX
SQ Sequence 32 AA;

Query Match 100.0%; Score 113; DB 21; Length 32;
Best Local Similarity 100.0%; Pred. No. 6.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPDINPAWYASRGIRPVGRF 20
||||||||||||||||||||
Db 12 tpdinpawyasrgirpvgrf 31

Search completed: April 17, 2001, 15:38:42
Job time: 349 sec